



# UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría en Filosofía de la Ciencia  
Instituto de Investigaciones Filosóficas  
Filosofía de la Ciencia

Análisis y evaluación de la posibilidad de extender el principio de caridad en la versión de Donald Davidson como herramienta para la evaluación de la racionalidad del comportamiento humano

## Tesis

QUE PARA OPTAR POR EL GRADO DE  
Maestro en Filosofía de la Ciencia

PRESENTA:  
Rodrigo Munguía Noriega

### DIRECTORES DE TESIS

Dr. Alejandro Vázquez Del Mercado Hernández  
Instituto de Investigaciones Filosóficas

Dra. Ángeles Eraña Lago  
Instituto de Investigaciones Filosóficas

Ciudad Universitaria, Cd. Mx, diciembre de 2020



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## Índice

Agradecimientos	4
Introducción	5
Capítulo 1: La Interpretación radical en Donald Davidson	9
1.1 Puesta en marcha de la teoría	11
1.1.2 Requisito de comprensión (RC)	12
1.1.3 Requisito de Evidencia (RE)	15
1.1.4 Restricción holística (HO)	16
1.2 El papel del principio de caridad en la interpretación radical	17
1.2.1 La intérprete	17
1.2.2 El Hablante	19
1.2.3 El primer esbozo del principio de caridad y su función dentro de la Interpretación radical	21
1.3 Conclusiones del capítulo	22
Capítulo 2: Análisis del Principio de caridad (PC) en Davidson	23
2.1 Reconstrucción y análisis del PC	24
2.1.1 Principio de Correspondencia (PCRR)	26
2.1.2 Principio de Coherencia (PCH)	28
2.1.3 De vuelta a la maximización	30
2.2 Las justificaciones del PC	33
2.2.1. La apelación a la necesidad del PC para la Interpretación radical	34
2.2.2 La apelación al Holismo o a la naturaleza de las creencias	37
Capítulo 3: Las extensiones del PC como herramienta para la evaluación de la racionalidad y una evaluación crítica de las mismas.	39
3.1 La propuesta de Stein sobre la extensión del PC a una herramienta para la evaluación de la racionalidad del comportamiento humano	40
3.1.1 El papel del PC	42
3.1.2 Extensión del PC para la evaluación de la racionalidad del comportamiento	44
3.1.3 Evaluación del PC dentro de la propuesta de Stein	47

3.2 La caridad como método de evaluación en las Ciencias Sociales: La versión de la extensión del PC en Thagard y Nisbett.	51
3.2.1 Los argumentos en contra de la aplicación del PC para la interpretación y su crítica.	53
3.2.2 Contra la extensión del PC para las conductas inferenciales	61
3.2.3 Contra la extensión del PC para la toma de decisiones.	64
3.3 Resumen y conclusiones sobre la extensión del PC.	66
4. Conclusiones generales y nuevas posibles rutas de indagación para el PC	68
4.1 Conclusiones generales	68
4.2 Otras rutas de indagación para el PC dentro del análisis de la racionalidad.	73
Bibliografía y Referencias	79

## **Agradecimientos**

Esta tesis fue realizada con la beca de maestría del Consejo Nacional de Ciencia y Tecnología por lo cual agradezco su apoyo durante mi estancia como estudiante de la Maestría en Filosofía de la Ciencia.

Agradezco a mi director y amigo Alejandro Vázquez del Mercado por haber confiado en mí, haber tomado mi proyecto y haberme orientado por el mejor camino para llevarlo a cabo. También agradezco a Ángeles Eraña por respaldar mi proyecto y ayudarme a llevarlo a término, a Jonatan García Campos por su dedicación a mi trabajo, por la lectura tan cuidadosa que hizo de éste y por haberme hecho comentarios y observaciones que realmente suscitaron dudas en mí que aún no puedo resolver; a Axel Barceló por los valiosos comentarios que me hicieron pensar mucho más allá de mi propio trabajo y me llevaron a reflexiones que lo fortalecieron, a Claudia Lorena García que con gusto aceptó ser mi sinodal y colaborar conmigo en esta tarea.

“Somos una decisión que alguien más tomó y que hoy se decide a sí misma” es mi definición de un ser humano. Agradezco a esas dos personas que tomaron esa primera decisión para que yo existiera y después la arrojaron con amor y esfuerzo para que decidiera mejor. Gracias, mamá; gracias, papá. Todo es una buena ocasión para agradecer.

## Introducción

¿Cómo descubrimos el significado que una persona atribuye a un nombre propio?

O de otra manera; si Carlos enuncia las siguientes oraciones:

1. César conquistó la Galia.
2. César cruzó el Rubicón.
3. César fue asesinado en los idus de marzo.
4. César era adicto al uso del ablativo absoluto.
5. César estaba casado con Boudica.

¿Cómo podemos saber qué significa “César” según el uso que Carlos hace de él?

Si suponemos que conocemos el significado de todas las palabras que Carlos utilizó, menos el de la palabra “César”, y tampoco sabemos de historia de Roma, al inicio; lo que nos queda hacer es unir todas las oraciones en conjunción y comenzar una investigación empírica sobre historia para ver quién cumple con todos los predicados que se adjudican a “César” y así llegar a conocer el significado que tiene para Carlos.

Una vez concluida nuestra investigación histórica, llegaríamos a la conclusión que hay un “César”, Julio César que satisface y, por ende, hace verdaderas a las primeras cuatro de las cinco oraciones que profirió Carlos, mientras que la quinta oración, ahora que sabemos mucho de historia, es probable que pueda corresponder a Prasutago, rey de los icenos.

Si lo que nos interesa, en realidad, es poder comprender a Carlos y comunicarnos con él, tenemos que tomar una decisión y apoyarnos en el Principio de Caridad para seleccionar al individuo, Julio César, que hace verdaderos a la mayor cantidad de enunciados y, por lo tanto, podríamos decir que pueden ser una buena traducción

de estos; Así nació el Principio de Caridad (PC) en el artículo del filósofo Niel L. Wilson, "Substance without Substrata" de 1959. La idea original del PC es lograr una traducción que maximice la verdad en la mayor cantidad de las oraciones cuyo significado buscamos conocer. La versión del PC que retoma Davidson directamente de Quine, y en el que se centrará este trabajo, además, de buscar maximizar la verdad del conjunto de oraciones que constituyen las creencias de un hablante, también asume, *prima facie*, la consistencia de dicho conjunto en la búsqueda de la comprensión de su lenguaje. Para Davidson el PC es fundamental para desarrollar su teoría de la Interpretación radical.

A lo largo de varias décadas, muchas y muchos autores han intentado extender el PC como una herramienta metodológica para evaluar la racionalidad del comportamiento humano. El presente trabajo se centra en una propuesta de extensión del PC que hace Edward Stein en su libro "Without Good Reason" (1996) y retoma y analiza los argumentos que presentan Paul Thagard y Richard E. Nisbett en su artículo "Rationality and Charity" para defender la tesis de que no es posible utilizar el PC como herramienta para la evaluación de la racionalidad del comportamiento humano. Sin embargo, presentaré al final algunas extensiones existentes hacia otras áreas.

Para demostrar mi tesis, comenzaré en el Capítulo 1 esclareciendo el papel que tiene el PC en la versión de la Interpretación radical de Donald Davidson; y con miras a ello, en 1.1) expondré y explicaré la teoría de la interpretación radical, en 1.2) haré una breve exposición del PC y de su papel dentro de dicha teoría, y en 1.3) extraeré algunas repercusiones y consecuencias de la Interpretación radical y del papel del PC dentro de ésta.

Posteriormente, en el Capítulo 2 mi objetivo será analizar y explicar a detalle el PC en Davidson para, más adelante, en el capítulo siguiente, evaluar la posibilidad de extenderlo como principio o herramienta de evaluación de la racionalidad, y para llegar a ello, en 2.1) ofreceré una recopilación de las versiones del PC, tomando como referencia los textos donde Davidson lo enuncia y las reconstrucciones de algunas de sus comentadoras y críticas para reconstruir y argumentar a favor de una versión que utilizaremos a lo largo de este trabajo, mientras que, en 2.2) analizaré algunas de las justificaciones que se han ofrecido del PC y que aparecen en la obra de Davidson.

En el Capítulo 3, mi objetivo será evaluar una propuesta y dos críticas que existen para extender el PC de un principio que sirve para la interpretación a una herramienta que evalúe la racionalidad del comportamiento humano. Con este fin, desarrollaré los siguientes puntos: 3.1) expondré y evaluaré el argumento de la propuesta de extensión de Edward Stein, en 3.2) llevaré a cabo una evaluación crítica de los argumentos en contra de la extensión del PC en la versión de Thagard y Nisbett, y en 3.3) elaboraré un breve resumen y conclusiones del capítulo. Se presentarán tanto argumentos para objetar las propuestas particulares para la extensión del PC, antes citadas, como un argumento para una posible objeción general.

Finalmente, una vez analizadas y rechazadas las propuestas de extensión del PC como herramienta para el análisis de la racionalidad humana, en el Capítulo 4 se ofrecerán algunas conclusiones generales en 4.1) y en 4.2) se presentarán dos rutas alternativas para utilizarlo en la elaboración de una teoría caritativa de la racionalidad y como principio para una comprensión del comportamiento dentro del

programa de investigación del Análisis Racional; un programa empírico cuyo objetivo es explicar la función y el propósito de los procesos cognitivos y cómo es que estos son adaptativos con respecto a su entorno. Una de sus implicaciones importantes para la filosofía consiste en que ofrece un modelo para comprender la relación entre los principios formales de racionalidad, como la probabilidad y la teoría de la decisión y la racionalidad cotidiana, es decir, la que utilizamos en la vida diaria (Chater y Oaksford, 2000).

## Capítulo 1: La Interpretación radical en Donald Davidson

The story I tell about how radical interpretation is possible should be viewed as an informal proof that the evidence assumed available plus the constraints on the structure of the pattern of a person's beliefs, values and language, suffice to yield an interpretive theory for understanding that person.  
Davidson, 1994

El objetivo del presente capítulo es esclarecer el papel que tiene el PC en la versión de la Interpretación radical de Donald Davidson; con este fin en 1.1) Expondré la teoría de la Interpretación Radical, en 1.2) Haré una breve exposición del PC y de su papel dentro de la teoría, y en 1.3) elaboraré algunas repercusiones y consecuencias de la Interpretación radical y del papel del PC dentro de ésta.

Es necesario explicar en qué consiste la interpretación radical porque tenemos que saber cómo funciona el PC dentro de ella y cuál es su rol dentro para descifrar el meollo del problema: cómo dar cuenta de las creencias de un hablante y de los significados de éstas.

Podemos decir que comprendemos una aserción bajo una cierta interpretación que hacemos de ella. En este sentido, interpretar no es nuevo para nadie, lo hacemos todo el tiempo, incluso con hablantes de la propia lengua. Sin embargo, una teoría de la interpretación radical sirve para los casos en los que buscamos interpretar a un hablante desde una evidencia distinta al conocimiento de su lengua, al de sus creencias o sus deseos; de ahí que, la pregunta de la que parte la teoría sea: ¿Qué

podríamos saber (no que ya sabemos) que nos ayude a interpretar los enunciados que produce un hablante?

Es importante comenzar diciendo que la teoría de la interpretación radical no fue elaborada con el fin de ofrecer un método empírico para ser utilizado en el trabajo de campo, sino que su sentido es arrojar luz a la investigación sobre la naturaleza de conceptos tales como el lenguaje, el significado, la referencia y las actitudes proposicionales, entre otros; específicamente, sobre la relación que existe entre el lenguaje y las intenciones, lo cual, además, revelará aspectos importantes de la comunicación humana.

Para Davidson, “significado y creencia desempeñan papeles interrelacionados y complementarios en la interpretación del discurso” (Davidson, 1974a, 151). Esta interrelación quiere decir que no es posible inferir una creencia sin conocer el significado lingüístico, al igual que no es posible inferir el significado lingüístico sin conocer las creencias. En la búsqueda de la salida de este círculo entre dos desconocidos, creencia y significado, es que cobra sentido la idea de una teoría radical de la interpretación.

La interdependencia de creencia y significado resulta evidente de esta manera: un hablante sostiene que una oración es verdadera debido a lo que la oración (en su lenguaje) significa y debido a lo que él cree. Sabiendo que él toma a la oración por verdadera y conociendo el significado podemos inferir su creencia; dada la suficiente información sobre sus creencias, podemos tal vez inferir el significado. (Davidson, 1973)

Nosotros podemos saber lo que un hablante quiere decir con una expresión cuando sabemos cuál es la creencia que intenta decirnos. Por tanto, identificar el significado es pues inseparable de la tarea de atribuir creencias. Sólo si tenemos acceso a las creencias del hablante, entonces, nos será posible identificar las instancias de las oraciones que él cree que son verdaderas; así, podemos hacer corresponder las creencias con las oraciones y de esa forma llegar a su significado.

### 1.1 Puesta en marcha de la teoría

Podemos decir, que la interpretación radical es un proceso global en el que la asignación de verdad a las emisiones y la asignación de estados mentales como creencias y deseos a un sujeto se llevan a cabo simultáneamente y se condicionan de manera recíproca. Davidson nos plantea ejemplos sobre situaciones en las que nos encontramos frente a un hablante de una lengua que desconocemos. Lo primero que habría que decir es que el procedimiento de la interpretación radical funciona como una idealización de una situación de interpretación, y en ese sentido, no es el procedimiento que utilizaría cualquier persona frente dicha situación, sino un procedimiento que requiere de una teórica radical como intérprete.

Para exponer el funcionamiento de la teoría retomaré la pregunta sobre qué podríamos saber o, en otros términos, cuáles son los requisitos que debe tener presentes la intérprete para poder comprender el lenguaje del hablante y los iremos explicando a la par de la teoría.

### 1.1.2 Requisito de comprensión (RC)

El primer paso frente a una situación hipotética en la que como intérpretes radicales nos encontramos ante la emisión de un hablante o un grupo de hablantes de una lengua desconocida es asignar contenido proposicional a aquello que logremos identificar como sus oraciones. En la medida en que nuestra asignación sea exitosa, podemos decir que las oraciones que nosotras formulamos como intérpretes proporcionan las condiciones de verdad de las aserciones que escuchamos y que este procedimiento nos sirve como base para la interpretación. En otras palabras, que iniciamos la elaboración de una teoría del significado para las oraciones presentes y futuras del hablante.

En un contexto de interpretación radical, Davidson sugiere que la intérprete construye una teoría del significado para el lenguaje de los hablantes partiendo de la construcción de una teoría tarskiana de la verdad (Davidson, 1967). En otras palabras, como intérpretes lo que hacemos es producir oraciones T (llamadas así por Tarski) con la forma de bicondicionales que tienen del lado izquierdo una oración del hablante de la que se predica verdad y del derecho una oración del intérprete. Suponiendo que el hablante esboza algo como: *Der Hund bellt* en un momento determinado Una oración T sería:

*“Der Hund bellt”*, (una oración en el lenguaje objeto) es verdadera en alemán en un momento determinado, sí y sólo sí, el perro está ladrando.

Para que la oración T sea verdadera, la oración del lado derecho debe tener las mismas condiciones de verdad que la oración del hablante citada del lado izquierdo. La teoría del significado para el lenguaje L será un número indeterminado de oraciones T que nos permitan realizar interpretaciones cada vez más precisas de las nuevas aserciones.

Como Davidson estipula, en la construcción de la teoría, la intérprete radical, primero trata de identificar los predicados, los términos singulares, los cuantificadores y la identidad para fijar la estructura lógica del lenguaje del hablante; luego se concentra en las oraciones sostenidas como verdaderas<sup>1</sup> y otras como falsas en función de cambios en el mundo que él deberá reconocer. Por último, se ocupa de las oraciones sobre las que no hay acuerdo entre los hablantes, en caso de que fueran varios, o cuyo valor de verdad no parece depender de cambios reconocibles en el mundo. El contenido del pensamiento y el lenguaje se determinan a través de un proceso triangular en el que están relacionados: la intérprete, el hablante y el mundo, es decir, el hablante emite una aserción con base en algo que observa en el mundo y la intérprete, que se encuentra en la misma circunstancia y contexto que el hablante, tiene la misión de desentrañar tanto el significado lingüístico como el contenido de la creencia a la que hace referencia la aserción. Así, tal como lo decía anteriormente, cuando la intérprete cuente con un conjunto suficiente de oraciones T, podrá evaluar el rol de los términos en las atribuciones de verdad de las oraciones; de ese modo, puede determinar su significado.

---

<sup>1</sup> La actitud de “sostener como verdadera” una emisión supone la atribución de intenciones al hablante.

¿Y bajo qué tipo de teoría semántica se pueden interpretar las oraciones? Para Davidson basta con tener una teoría que provea de una definición de verdad y que a partir de ésta se pueda extraer el significado. En palabras de Davidson: “una teoría del significado para un lenguaje L muestra cómo los significados de las oraciones dependen de los significados de las palabras si contiene una definición (recursiva) de verdad para L” (Davidson, 1967) y con respecto al concepto de verdad, para Davidson, decir que una oración “O” es verdadera no es otra cosa que decir que O, es decir, sólo es necesario desentrecomillarla, dicho en la jerga davidsoniana, ya que sólo es otra forma de afirmar la oración. Esta tesis es la base de la definición semántica de Tarski del concepto de verdad. Davidson aplica esta definición a los lenguajes naturales, aunque Tarski niega que se la pueda aplicar, y la toma como la base del desarrollo de una teoría del significado, con la salvedad de que mientras Tarski utiliza el significado para definir la verdad, Davidson invierte esta relación y utiliza la verdad para llegar al significado, ya que él toma el concepto de verdad como primitivo.

Una teoría que utiliza las condiciones de verdad del bicondicional para interpretar las aserciones de un hablante satisface lo que Davidson considera el primer requisito para la interpretación al que llamaremos aquí el Requisito de Comprensión (RC) una teoría que a partir de elementos finitos ayude al intérprete a comprender cualquier oración que un hablante pudiera emitir.

### 1.1.3 Requisito de Evidencia (RE)

Por otra parte, ¿qué tipo de evidencia empírica podría apoyar a una teoría del significado como la anterior? La base sobre la que se asigna el contenido a las emisiones del hablante son las circunstancias y el contexto en el que se lleva a cabo la emisión, es decir, observando detenidamente, en varias ocasiones podemos intentar descubrir la actitud y el contenido proposicional de un hablante poniendo atención a lo que ocurre a su alrededor al momento en el que se profiere una emisión. Davidson (1973) supone que, con la debida agudeza y perseverancia, somos (como intérpretes radicales) capaces de descubrir actitudes como una creencia, algo que sostiene como verdadero; un deseo, algo que quiere que sea verdadero o una preferencia, algo que se valora por encima de otra cosa, de un hablante frente a un objeto en un momento específico, incluso si no conocemos el significado de su oración.

Una vez que reconocemos en varios hablantes un grupo de oraciones podemos llegar a decir que, en un momento determinado y bajo ciertas circunstancias particulares, sostienen como verdaderas y elaboramos una oración T como:

*“Der Hund bellt”* (el lenguaje objeto) es verdadera en alemán para un hablante en un momento determinado, sí y sólo sí, el perro está ladrando cerca de dicho hablante en un momento determinado.

Un ejemplo de lo que cuenta como evidencia de eso es que nuestro hablante, Hans, perteneciente al grupo de personas que hablan alemán, se encuentre en una situación particular, supongamos una tarde frente al portón de una casa en donde habita un poodle, un perro que suele ladrar mucho; que el perro comience a ladrar y que Hans, entonces, profiera la oración *Der Hund bellt*. Nuestra intérprete, que se encuentra en la misma circunstancia y contexto que Hans, escucha la oración que emite y la interpreta utilizando una oración T como la anterior. De este modo, tenemos un segundo requisito para la interpretación cubierto, el que denominaremos, el Requisito de Evidencia (RE), y que se define como los datos que encontramos en el mundo al momento de la emisión de una oración por parte de un hablante bajo ciertas circunstancias y bajo cierto contexto que permiten a la intérprete inferir las oraciones T necesarias para la construcción de la teoría del significado del lenguaje L.

#### 1.1.4 Restricción holística (HO)

Hasta el momento, RC nos garantiza que construimos una teoría del significado con la que somos capaces de comprender cualesquiera de las oraciones que el hablante de un lenguaje L pudiera esbozar. RE nos aporta evidencia necesaria para la posibilidad de generar en el lado derecho del bicondicional de las oraciones T, oraciones que se restrinjan al contexto y circunstancias de emisión bajo las cuáles podamos considerar qué es lo que el hablante sostiene como verdadero. Sin embargo, aún queda abierta la posibilidad de que la intérprete formule una o varias oraciones T donde algo como “él árbol es parte de la naturaleza” es verdadera, si y

sólo si, el cielo es azul o un soltero es un no casado compartan valores de verdad y, por lo tanto, se esté tentado a decir que el lado derecho es una interpretación del lado izquierdo, lo cuál sería aberrante.

Davidson recurre a una restricción extra (Davidson, 1973 y 1974a); para garantizar que el lado derecho del bicondicional de una oración T tenga las mismas condiciones de verdad del lado izquierdo debemos apelar a que no sólo tenemos una oración T, sino a que sabemos que la teoría es correcta y tenemos un conjunto grande de oraciones T consistentes con la traducción que hacemos de una nueva oración por parte del hablante; esto es una Restricción Holística (HO) que por medio de la coherencia entre las oraciones T de la teoría se garantiza la verdad del resto de las oraciones que traducen a cada nueva aserción de un hablante.

## 1.2 El papel del principio de caridad en la interpretación radical

### 1.2.1 La intérprete

En la búsqueda de las oraciones T, es necesario que la intérprete presuponga: 1) que los hablantes emiten aserciones, es decir, dicen algo que puede ser verdadero o falso acerca del mundo, 2) dichas aserciones reflejan sus creencias, y 3) que la mayoría de éstas son correctas, en este sentido, se presupone la racionalidad del hablante, y esto será importante tomarlo en cuenta en el desarrollo posterior de este trabajo.

En el periodo de construcción de la teoría, la intérprete pasa por alto de manera deliberada la posibilidad de error de los hablantes. Cuando ha recabado suficiente información del asentimiento y disentimiento de los hablantes frente a distintas aseveraciones, la intérprete intentará predecir la reacción del hablante frente a una nueva oración. Si el resultado no es el esperado, la intérprete puede atribuir el error a la teoría, a sus propias creencias o, incluso, pero como último caso, a las creencias del hablante. Esto último no lo puede hacer al comienzo de la construcción de la teoría. La intérprete debe suponer que el hablante asevera correctamente, y la intención es que los contenidos de las creencias del hablante se vayan descubriendo en la medida en que la teoría resulte exitosa. Es importante para el desarrollo de este trabajo destacar que Davidson no blinda la posibilidad de atribuir un comportamiento lingüístico ni doxástico irracional a un hablante:

El consejo metodológico de interpretar en una forma que optimice el acuerdo no debe concebirse apoyado en una presunción caritativa acerca de la inteligencia humana que podría resultar falsa. Si no podemos encontrar una forma de interpretar que las emisiones y otras manifestaciones de la conducta de una creatura son reveladoras de un conjunto de creencias ampliamente consistentes y verdaderas según nuestros propios patrones, no tenemos razones para considerar que esa creatura es racional, tiene creencias o dice algo. (Davidson, 1973)

Para Davidson ser racional consiste en tener actitudes proposicionales como creencias, deseos, las intenciones, etcétera. Un gran número de creencias cohesionadas y consistentes y, además, en su mayoría verdaderas son suficientes para atribuir racionalidad a cualquier animal. Dado que, por naturaleza, actitudes proposicionales se dan en racimos, tener un gran número de creencias consistentes también supone que ser racional está relacionado con la capacidad para realizar las inferencias necesarias para mantener dicha consistencia.

Hasta aquí los deberes y los requisitos para la aplicación del PC por parte del intérprete. Sin embargo, también existen algunos constreñimientos que debe seguir un hablante para el éxito del funcionamiento de la caridad y, posteriormente, de la interpretación.

### 1.2.2 El Hablante

Davidson no cree que los hablantes forzosamente deban mantener las mismas convenciones de habla que el resto de los miembros de la comunidad, ya que nadie puede ni debe estar obligada con un lenguaje, aunque no es poco frecuente que los hablantes mantengan usos y comportamientos lingüísticos muy parecidos a los que tiene su comunidad; sin embargo, lo que sí es posible afirmar, de manera general, es que cualquier persona sigue al menos un propósito al comunicarse: ser comprendidas como esperamos para conseguir aquello que se haya propuesto.

El propósito del hablante funda el parecido lingüístico con la comunidad y, al mismo tiempo, una norma que posibilita el que se pueda atribuir el error lingüístico, a pesar de que el uso de una palabra pueda ser significativo sin que necesariamente concuerde con los usos que le da la comunidad de manera cotidiana.

En este sentido, un deber del hablante es seguir su intención lingüística, ciertamente no siempre tenemos la intención de aseverar cosas, a veces llevamos a cabo preguntas, imperativos o exclamaciones, mientras que, en algunos otros casos, la intención puede ser aseverar algo; en estos casos, el hablante se compromete a

levantar su emisión si reconoce que es falsa. Su intención de aseverar, literalmente decir algo que sea verdadero, lo compromete a reconocer que ha errado si se le muestra que lo que ha dicho es falso.

Por otra parte, el hablante se compromete a usar los términos del mismo modo en el futuro. Una aserción establece un vínculo entre las palabras que la conforman y ciertos significados que ellas expresan y que el hablante deberá respetar en sus emisiones. Un requisito para el éxito de la construcción de una teoría del significado para el lenguaje de un hablante es que éste mantenga estable el modo en que utiliza los términos ya que, si él varía el uso de sus términos, cuando la intérprete intente recrear el rol estructural de los términos a partir de las oraciones T para descifrar su significado, se encontrará con varias inconsistencias que dificultarán la tarea a realizar y sus predicciones fallarán en una enorme cantidad de casos. Por otra parte, si el hablante no es constante en el significado que atribuye a los términos, la intérprete no podrá reconocer las oraciones que son verdaderas en función de los cambios en el mundo que debe reconocer. Si, por ejemplo, el hablante refiere con la palabra “noche” a veces al día y a veces a la noche, la intérprete no podrá reconocer que la verdad de la oración del hablante “ahora es de noche” depende de cambios en el mundo.

### 1.2.3 El primer esbozo del principio de caridad y su función dentro de la Interpretación radical

La finalidad del principio de caridad es la posibilidad de la comprensión lingüística a través de la presuposición de la racionalidad del hablante como poseedor de creencias mayoritariamente verdaderas.

Y, desde luego, el hecho de que una teoría no erija a los hablantes en poseedores universales de verdades no representa una adecuación de la teoría; la meta no persigue el absurdo de hacer desaparecer el disenso y el error. La idea es más bien que solo el acuerdo amplio da el marco dentro del cual pueden interpretarse disputas y errores. La tarea de dar un sentido a las emisiones y conductas de los otros, aún a su conducta más aberrante, nos exige encontrar una gran dosis de razón y de verdad en ellos. Ver un exceso de sinrazón en los otros es simplemente socavar nuestra capacidad de comprender a qué se debe que sean tan poco razonables. (Davidson, 1974a)

Así, podemos decir como un primer acercamiento que se pretende refinar en la siguiente sección que el principio de caridad para Davidson es una herramienta que sirve al proceso de interpretación radical para atribuir y fijar la verdad provisional al sistema de creencias de un hablante, o una comunidad de hablantes, en la búsqueda de la optimización del acuerdo y presuponiendo la racionalidad de éste o de dicha comunidad, con el fin de lograr la comprensión de su lenguaje.

La estrategia es fijar la creencia para lograr obtener el significado asignando condiciones de verdad según lo que la intérprete considere verdadero, en ese sentido, si la asignación no resulta consistente es ella quien debe ajustar la teoría en la búsqueda de maximizar el acuerdo con el fin de lograr la comprensión.

De acuerdo con lo anterior, el papel del PC en la Interpretación Radical es doble: hacer que los datos que fueron recabados a partir de las circunstancias observables convertidos en oraciones estén disponibles como evidencia para formar una teoría de oraciones T. Mientras que, por otro lado, evalúa y jerarquiza las teorías formadas de oraciones T para que la intérprete radical elija la que más se ajuste a los datos que posee. La teoría de oraciones T que mejor se ajuste a los datos obtenidos del hablante, es la teoría más caritativa.

### 1.3 Conclusiones del capítulo

A grandes rasgos, las reflexiones semánticas de Davidson giran en torno a la pregunta por la posibilidad de la comprensión y la comunicación lingüística. La estrategia de la interpretación radical es la siguiente: tomar las aseveraciones del hablante, fijar sus creencias presuponiendo y conservando provisionalmente la verdad de éstas (principio de caridad), indagar empíricamente la verdad a través del asentimiento y disentimiento del hablante, con ello, recrear el significado y hacer esto de manera constante hasta refinar lo más posible el significado para corroborar las creencias.

La interpretación radical delimita una forma distinta a la de otras tradiciones de ver el lenguaje y el significado. Davidson considera que es erróneo entender algo como un lenguaje en abstracto, al igual que los significados de las palabras independientemente de un acto particular de interpretación y una teoría del

significado; es decir, no existe un lenguaje en abstracto independientemente de un hablante y un intérprete concretos en un ejercicio de comprensión verbal.

Por otra parte, otra de las consecuencias que busca mostrar la interpretación radical es la naturaleza del lenguaje es social y el significado es público y está al alcance de cualquier intérprete. Es decir, una de las características principales del lenguaje es su “interpretabilidad” y gracias a ella se puede llegar tanto al significado como a la posibilidad de determinar intenciones, al menos, las creencias, y en este sentido, Davidson intenta fundar la validez de la interpretación radical en que, de hecho, somos capaces de entendernos y conocer mutuamente nuestras creencias. Por lo tanto, la teoría es, al menos, plausible.

## **Capítulo 2: Análisis del Principio de caridad (PC) en Davidson**

El objetivo del capítulo es analizar y explicar a detalle el PC en Davidson para posteriormente, en el capítulo siguiente, evaluar la posibilidad de extenderlo como principio o herramienta de evaluación de la racionalidad para llegar a ello, desarrollaremos los siguientes puntos: 2.1) ofreceré un análisis más detallado del PC, tomando como referencia los textos donde Davidson lo enuncia y las versiones de algunas de sus comentadoras y críticas; 2.2) analizaré algunas de las justificaciones que se han ofrecido del PC tanto por parte de Davidson como de sus comentadoras.

## 2.1 Reconstrucción y análisis del PC

A pesar de lo central del papel que juega dentro de su filosofía del lenguaje, Davidson no formula ni desarrolla una explicación canónica estructurada de su versión del Principio de Caridad como tal, lo cual ha dado pie a múltiples y diversas reconstrucciones por parte de sus comentadoras y críticas<sup>2</sup>. Desde su primera enunciación de éste en su artículo de 1967, Verdad y significado, las menciones y versiones del PC se encuentran desperdigadas a lo largo de una buena parte de su obra.

En Verdad y significado, Davidson escribe:

El lingüista intentará luego construir una caracterización de la verdad para el extranjero que produzca, en lo posible, un mapeo de las oraciones consideradas verdaderas (o falsas) por el extranjero con las oraciones consideradas verdaderas (o falsas) por el lingüista. Suponiendo que no se encuentra una coincidencia perfecta, el residuo de oraciones consideradas verdaderas traducidas por oraciones consideradas falsas (o viceversa) es el margen de error (foráneo o doméstico). Cierta caridad en las palabras y los pensamientos de los otros es inevitable también en otra dirección: así como debemos maximizar el acuerdo o arriesgarnos a no encontrar sentido a lo que el extranjero está diciendo, también debemos maximizar la auto consistencia que le atribuimos bajo pena de no entenderlo *a él*. (Davidson, 1967)

Haciendo énfasis en que el PC tiene una misión maximizadora de dos elementos, por un lado, con él se pretende maximizar el acuerdo, es decir, la confluencia entre las creencias del hablante y las que tiene el lingüista y, por otro lado, la consistencia del hablante. Maximizar la consistencia quiere decir interpretar las oraciones que el

---

<sup>2</sup> Glüer, Lepore, Ludwig, Goldberg, Pagin, por citar solo algunos de los que analizaremos en este trabajo.

hablante sostiene como verdaderas de manera que respeten los principios de la Lógica. Sin embargo, hasta aquí no menciona la forma ni el elemento a partir del cual se puede maximizar el acuerdo.

En el texto de "Interpretación Radical" de 1973, Davidson escribe una de las formulaciones que mejor hacen explícita la función que tiene el PC en el proceso de interpretación:

Este método pretende resolver el problema de la interdependencia entre creencia y significado tomando la creencia como constante hasta donde sea posible en tanto se resuelve el significado. Esto se consigue mediante la asignación de condiciones de verdad a oraciones extranjeras que hacen que los hablantes nativos tengan razón cuando es plausiblemente posible, por supuesto según nuestro propio punto de vista de lo que es correcto. (Davidson, 1973)

La forma de maximizar el acuerdo consiste, entonces, en intentar interpretar las oraciones de los hablantes, de manera que éstas coincidan en su mayoría con lo que la intérprete considera verdadero. En ese sentido, el acuerdo consiste en que lo que es verdadero para el hablante lo sea para la intérprete también.

En un texto posterior, Tres variedades de conocimiento, Davidson pone nombre a los dos aspectos que debe maximizar el intérprete y que son también principios constitutivos del PC: El Principio de Correspondencia (PCRR) y el Principio de Coherencia (PCH) los cuales revisaremos más a detalle por separado.

El proceso de separar el significado y opinión recurre a dos principios clave que deben aplicarse si un hablante puede ser interpretado: El Principio de Coherencia y el Principio de Correspondencia [...] Ambos principios pueden denominarse principios de caridad (y se los ha denominado así), un principio dota al hablante de una módica cantidad de lógica y el otro le dota en un cierto grado con lo que el intérprete considera que son creencias verdaderas acerca del mundo. (Davidson, 1991)

Veamos ambos principios más a detalle.

### 2.1.1 Principio de Correspondencia (PCRR)

Al respecto, Davidson dice:

El Principio de Coherencia lleva al intérprete a considerar que el hablante está respondiendo a los mismos rasgos del mundo a los que él (intérprete) respondería en circunstancias similares. (Davidson, 1991, 188)

La correspondencia es la forma de anclar con el mundo los contenidos de las creencias del hablante que el intérprete fija en una fase inicial y a las cuáles asigna, poco a poco, condiciones de verdad con base en lo que él mismo cree que es verdadero.

El fundamento del anclaje de los contenidos que garantiza que la creencia está en contacto con el mundo es que, por un lado, Davidson considera que las creencias están relacionadas en un entramado que sostiene su adecuación al resto de las creencias por medio de la coherencia; sin embargo, y más relacionado con la correspondencia, Davidson considera que son los objetos del mundo los que causan nuestras creencias.

Lo que impide el paso a un escepticismo global acerca de los sentidos es, en mi opinión, el hecho de que debemos en los casos más sencillos y metodológicamente más básicos considerar a los objetos de una creencia como la causa de esa creencia (Davidson, 1983, 212).

De modo que, por un lado, las creencias vienen en racimos entrelazados de manera coherente y, por otro lado, al menos el contenido de nuestras creencias perceptuales es producido causalmente por los objetos que observamos.

Lo anterior, para Davidson, garantiza que los contenidos que el hablante observa y que se encuentran presentes en las oraciones que emite sean los mismos a los que la intérprete atiende cuando se presenta una emisión en un contexto determinado, pero surge el problema de que, aun concediendo que el hablante y la intérprete coincidan o estén de acuerdo con los contenidos, la verdad de las oraciones que el hablante sostiene como verdaderas, queda totalmente a juicio de la intérprete. Es decir, es la adecuación a las creencias de la intérprete lo que se va a tomar por verdadero, de modo que bien podría ser que ésta hubiera generado la interpretación de un conjunto de oraciones coherentes y con las que coincide con el hablante, pero que estas oraciones no fueran verdaderas.

Desde luego, la anterior es una objeción plausible que no corresponde a este trabajo, pero en último término, para Davidson lo más importante es garantizar la interpretabilidad y la comunicabilidad de nuestras creencias. La garantía de la comunicación está en una triada (Davidson, 1991) que se establece entre dos sujetos que intercambian sus creencias y un mundo que sirve como garante de que

lo que ambos son incitados a creer es causado por las mismas características del mundo si se encuentran bajo las mismas circunstancias y contexto y captado perceptualmente por ellos, al igual que lo es lo que mutuamente emiten el uno hacia el otro.

Por último, debido a que, como vimos, para Davidson el contenido de las creencias perceptuales es causado por el mundo y en las mismas circunstancias de observación de ser compartido tanto por el hablante como por el intérprete, lo cual garantiza el acuerdo, el papel de las oraciones de observación o las que tienen elementos perceptuales más cercanos a ambos juega un papel muy importante en la constitución de la teoría en pro de maximizar el acuerdo entre ambos y ofrecer coherencia al sistema en construcción.

### 2.1.2 Principio de Coherencia (PCH)

En Tres variedades de conocimiento, Davidson escribe: “El Principio de Coherencia lleva al intérprete a descubrir un cierto grado de consistencia lógica en el pensamiento del hablante” (Davidson, 1991, 288). Dado que identificamos que nuestras creencias están íntimamente relacionadas unas con otras: mi creencia de que la comida de la calle al aire libre y en zonas muy concurridas puede ser más susceptible de recibir virus o bacterias que se encuentran en el aire, está relacionada con mi creencia de que podría enfermarme si como en la calle y, al mismo tiempo, esta creencia está relacionada con mi creencia de que es mejor comer en mi casa, entre otras miles de creencias más, de modo que cuando le

adscribo a un hablante la creencia de que no es bueno comer alimentos que se encuentran expuestos por mucho tiempo al aire libre en zonas concurridas, entiendo que puedo adscribir a esta persona muchas otras creencias relacionadas con ella y esto mismo aplica para todas las creencias que puedo interpretar para un hablante.

Las creencias están relacionadas de un modo inferencial, una parece seguirse de la otra o una podría ser o parecer una razón para creer la otra y así sucesivamente. De modo que es este tipo de pegamento lógico lo que articula la coherencia y ofrece la consistencia de todos los conjuntos o racimos de creencias, y al mismo tiempo, es también lo que conecta al PCH con la racionalidad. En palabras de Davidson: “La interpretación con éxito inviste necesariamente a la persona interpretada de una racionalidad básica”. (Davidson, 1991, 288).

A estas alturas, entonces, es posible notar dos características de las creencias: que vienen en conjuntos o en racimos (no es posible tener una creencia aislada) y estos están relacionados lógicamente; es decir: son conjuntos consistentes. Sin embargo, cabe adelantar aquí que esto no compromete a la intérprete a no atribuir inconsistencia en las creencias del hablante frente a evidencia flagrante. Davidson considera esta característica como un holismo y, al mismo tiempo, que es una característica de todo aquello que pertenezca a lo mental. Tomando en consideración estas características es que se puede afirmar que el PCH consiste en la determinación holística de los significados a partir de la actitud de sostener como verdaderas a las creencias por parte de los hablantes, lo cual determina las creencias y los contenidos de éstas.

Por último, podemos decir que el PCH inviste de una forma de racionalidad básica que permite interpretar bajo la coherencia las creencias del hablante. Sin embargo, por la naturaleza misma de las creencias, venir en conjuntos relacionados lógicamente, el sólo hecho de poder adscribir una creencia a un agente, supone que dicho agente es racional. Si se aceptan estas características, entonces, no es extraño dar cuenta de por qué Davidson sostiene que no sólo no es difícil lograr maximizar el acuerdo, uno de los objetivos del PC, sino que es relativamente sencillo ver que, dadas estas restricciones, cuando tenemos o conocemos una creencia de una persona y su contenido estamos a un paso de conocer también un gran conjunto de sus creencias. Y si a esto le agregamos que sus creencias las conocemos de acuerdo con lo que nosotros consideramos verdadero, el acuerdo entre la intérprete y el hablante es casi un hecho. No hay que olvidar en este punto que la intérprete no proyecta sus creencias en el hablante a su capricho, lo cual podría dar pie a sospechas sobre casos patológicos del intérprete, sino que Davidson cree que las creencias son causadas por el mundo que comparten intérprete y hablante.

### 2.1.3 De vuelta a la maximización

El principio de caridad no excluye la posibilidad ni de irracionalidad ni de inconsistencia. Tal como lo vimos, se parte desde el PCH de considerar *prima facie* que el hablante sostiene oraciones como verdaderas y que éstas son creencias que posee, y que, en este sentido, por la naturaleza de las creencias, éstas vienen en

conjuntos consistentes y que, por lo tanto, el hablante tiene un cierto grado de racionalidad que le permite tal consistencia en lo que cree.

Sin embargo, nada de esto excluye que se puedan tomar decisiones metodológicas como que la intérprete atribuya una creencia falsa al hablante en pos de salvaguardar la coherencia y la consistencia del conjunto que ya se han atribuido. En ese sentido, la interacción entre el PCRR y el PCH no necesariamente tiene que ser simétrica, una decisión que se tome para uno no necesariamente implica una acción para con el otro principio, aunque necesariamente se verá afectado como en el ejemplo anterior. Por otra parte, el tratar de hacer que las oraciones del hablante sean verdaderas en la medida de lo posible no es otra cosa que hacer que éstas correspondan con lo que la intérprete considera como verdadero, y esto que considera verdadero es a lo que se refiere Davidson con la labor del PC de maximizar el acuerdo: considerar que el hablante está en lo correcto cuando sea plausible que así sea.

De acuerdo con Davidson:

Queremos una teoría que satisfaga las restricciones formales de una teoría de la verdad y que maximice el acuerdo... en el sentido en que... (otros) estén en lo cierto, hasta donde nosotros podamos discernirlo, tan a menudo como sea posible. (Davidson, 1973, 147).

Más adelante, Davidson aclara que la idea de “maximizar” es ambigua, dado que nos es imposible contar las creencias, es más pertinente hablar de optimizar, en lugar de maximizar, dicho de otro modo, maximizar en último término quiere decir optimizar el acuerdo: “El precepto metodológico básico es, entonces, que una buena

teoría de la interpretación maximice el acuerdo, O dado que el número de oraciones es infinito, y dadas otras consideraciones por venir, una palabra mejor sería *optimizar*" (Davidson, 1975, 177)

La idea de optimizar el acuerdo significa, como lo habíamos mencionado, que existen diferentes tipos de acuerdo o creencias sobre las que, una vez más, metodológicamente, es mejor estar de acuerdo que en otras con el fin de lograr una teoría más eficaz para comprender mejor. Por ejemplo, el acuerdo sobre las creencias perceptuales o las que están más cerca de lo que se puede observar directamente es más importante que el acuerdo sobre creencias con respecto a conceptos teóricos o religiosos, si fuera el caso. A esto se refiere la optimización.

Los principios generales son relativamente fáciles de enunciar: el acuerdo sobre las leyes y las regularidades suele ser más importante que el acuerdo sobre los casos; el acuerdo sobre lo que es abierta y públicamente observable se favorece por encima del acuerdo sobre lo que se oculta, se infiere o pobremente observado; Las relaciones probatorias deben conservarse cuanto más se acerquen a ser constitutivas del significado.<sup>3</sup> (Davidson 1980, 157).

De este modo, la idea es que cierto tipo de creencias tienen prioridad interpretativa con respecto a otras en la conformación de una teoría del significado.

---

<sup>3</sup> La traducción es mía: "General principles are relatively easy to state: agreement on laws and regularities usually matters more than agreement on cases; agreement on what is openly and publicly observable is more to be favored than agreement on what is hidden, inferred, or ill observed; evidential relations should be preserved the more they verge on being constitutive of meaning".

## 2.2 Las justificaciones del PC

El objetivo del presente apartado es revisar algunas justificaciones posibles del PC con el fin de analizar posteriormente si éstas se sostienen de la misma forma en el caso en que extendamos la caridad al análisis de la racionalidad. Expondré dos justificaciones ofrecidas por Davidson, aunque cabe destacar que existen varias más que han sido elaboradas y desarrolladas por sus comentaristas y críticos. Más adelante diré un poco más sobre ellas.

Antes de comenzar, hay que decir que en varios pasajes y de diversas formas, Davidson deja claro que el PC es un principio necesario para la interpretación radical, pero cuya naturaleza exige que su justificación sea *a priori*:

Lo que hace la interpretación posible, entonces, es que podemos descartar *a priori* la idea de un error generalizado (Davidson, 1975, 176)

Un error generalizado supondría una atribución no caritativa de la verdad de las creencias del hablante por parte del intérprete, es decir, que sólo podemos saber que no puede haber un error generalizado porque asumimos por principio que quien tiene creencias es un ser racional, que las creencias mismas, por definición, deben ser consistentes y que el intérprete atribuye verdad a las creencias del hablante.

En "Sucesos mentales", Davidson escribe:

No puede considerarse como mera caridad conceder a la gente un alto grado de consistencia; es inevitable si queremos estar en posición de acusarla significativamente de error y de algún grado de irracionalidad. La confusión global, como el error total, es impensable no porque la imaginación vacile, sino porque demasiada confusión no deja nada

acerca de lo cual hallarse confuso, y el error masivo corroe el fundamento de la creencia verdadera, que es el único que nos permite explicar el fracaso. (Davidson, 1970, 280)

De este modo, Davidson supone que la caridad nos es impuesta en la búsqueda de la comprensión y, en último término, de la comunicación.

Y en *La psicología como filosofía*:

Del mismo modo como la satisfacción de las condiciones para la medición de longitud o de masa puede verse como algo constitutivo del rango de aplicación de las ciencias que emplean estas medidas, así la satisfacción de las condiciones de consistencia y coherencia racional pueden verse como cosas constitutivas del rango de aplicación de los conceptos tales como los de creencia, deseo, intención o acción. (Davidson 1974c, 298)

Haciendo una analogía entre las condiciones de medición para ciencias como la física y el PC como condición de la investigación de los conceptos intencionales. Es decir, apelando a leer el PC como una condición a priori.

### 2.2.1. La apelación a la necesidad del PC para la Interpretación radical

En varios pasajes, Davidson afirma que, además de indispensable, el PC es el único método posible para la interpretación radical.

El principio de caridad tiene un papel crucial en el método de Quine y un papel aún más crucial en mi variante del mismo. En cualquiera de los dos casos, el principio dirige al intérprete a traducir o interpretar de modo que lea sus propias normas o medidas para la verdad en la pauta de las oraciones que el hablante considera verdaderas. (Davidson, 1983, 209)

Y un poco más adelante:

La razón es que el único método – y por lo tanto irrecusable – de que puede disponer el intérprete pone automáticamente las creencias del hablante de acuerdo con las normas y

medidas de la lógica del intérprete, y por tanto otorga al hablante las verdades sencillas de la lógica. (Davidson, 1983, 210)

Este argumento trata de mostrar la necesidad de aceptar el PC como verdadero con base en el éxito de la Interpretación radical, aunque cabe adelantar, como lo aclararemos más adelante que esta línea es controversial y ha sido discutida por algunos autores como Lepore y Ludwig (2007), quienes incluso presentan alguna opción además del PC para la interpretación radical.

Ésta forma de presentar el PC es un típico caso de lo que en la tradición filosófica se ha dado por llamar: un argumento trascendental. Esta clase de argumentos se han utilizado tradicionalmente para combatir al escéptico en ciertas regiones del conocimiento. Me permitiré explicar su funcionamiento a partir de la distinción, también tradicional en filosofía entre condiciones necesarias, aquellas sin las cuáles algo no es verdadero, y condiciones suficientes, aquellas cuya verdad implica que algo sea verdadero, pero cuya falsedad, no necesariamente que aquello sea falso.

En este sentido, la forma de un argumento trascendental dice que, dado que X es condición necesaria para Y. Si sabemos que Y es verdadero, podemos concluir que X también tiene que serlo. Utilizando la versión del esqueleto lógico que ofrece Isabel Cabrera en la introducción del libro de “Argumentos trascendentales” (Cabrera, 1999, p. 8), el argumento quedaría de la siguiente manera:

Primero se afirma que algo es verdadero, en este caso, la afirmación es: “la interpretación radical funciona” la cual representamos como A, y después decimos que, si no se dieran ciertas condiciones de posibilidad C, en este caso, el Principio

de Caridad, entonces, A no sería posible; para así concluir la necesidad *a priori* de C, de la siguiente forma:

P1: A  
P2: ( $\neg C \supset \neg \Diamond A$ )  
 $\therefore \Box C$

Traducido al argumento para demostrar la justificación *a priori* del PC, el argumento quedaría así:

P1: La interpretación radical funciona  
P2. Sin el Principio de Caridad, no sería posible que la Interpretación radical funcionara.  
Por lo tanto, el Principio de Caridad es necesario.

Este argumento permitiría mostrar a Davidson que el PC es verdadero *a priori*. Aunque aún habría que mostrar la verdad de las premisas antes de aceptar la conclusión.

Con respecto a P1, Davidson atribuye a la interpretabilidad el fundamento de la comprensión entre los seres humanos, lo cual asume como un hecho que no requiere prueba. En el caso de P2, lo que habría que mostrar es, en primer lugar, que el PC es el único camino que puede tomar la interpretación radical, es decir, que no existe otra forma de descubrir sin ningún otro presupuesto el significado y las creencias del hablante al mismo tiempo, que no sea a través del PC. Con respecto a esta posible objeción, Lepore y Ludwig (2007) consideran que ciertamente el PC es el camino más simple, lo cual no implica que sea el mejor y otras autoras como Glüer, (2011) argumentan que el PC no es necesario, sino sólo

suficiente para la Interpretación radical, ya que existen otros caminos para ésta y, por lo tanto, si ella tiene razón el argumento es claramente falaz.

### 2.2.2 La apelación al Holismo o a la naturaleza de las creencias

De acuerdo con Davidson, tal como lo hemos visto antes, las creencias, por definición, no vienen solas, sino que las encontramos en conjuntos o en racimos lógicamente relacionados, aunque no hace explícita la relación lógica que las une, Kathrin Glüer considera que se puede inferir que las creencias van siendo una razón para las demás creencias:

Davidson escribe con mayor frecuencia como si las creencias mismas fueran razones para las creencias, pero eso es un poco engañoso. Que creo que  $p$  casi nunca es una (buena) razón para creer  $q$ . Más bien, las relaciones de razones se mantienen entre creencias porque lo que uno cree, es decir, el contenido de una creencia es una razón para una creencia más. Por tanto, podría existir una relación de razones entre la creencia de que  $p$  y la creencia de que  $q$ , si  $p$  es una razón para creer en  $q$ . (Glüer, 2011, 117)

De modo que, al poseer una creencia, uno además sostiene un conjunto de creencias que están relacionadas con ésta. Por ejemplo: si uno piensa que el papel se puede quemar fácilmente con el fuego, seguramente también piensa que una servilleta, un libro, o un cuaderno también se quemarían fácilmente; o si alguien tiene la creencia de que el COVID es un virus que se transmite a través de partículas suspendidas en el aire, también cree que es riesgoso estar muy cerca de una persona contagiada, cree que se deben evitar los lugares encerrados con mucha

gente y que debe apartarse de una determinada distancia de una persona que tose o estornuda, entre otras.

Davidson considera que el sólo hecho de tener un concepto, supone tener muchas creencias sobre dicho objeto o no se tiene dicho concepto. De esta forma, tener el concepto pasto, implica creencias como planta, que es de color verde, que se puede podar cada cierto tiempo, que se utiliza para practicar varios deportes, entre otras creencias.

De modo que, en resumen, el argumento holista a favor del PC se sustenta en la naturaleza de la creencia vista siempre como un conjunto coherente y consistente de creencias e invistiendo a la poseedora de las creencias de un cierto grado de racionalidad. Así que, si una persona tiene una creencia que corresponde con cómo se presenta el mundo en el momento  $t$ , tal como, que el viento levanta las hojas, y la intérprete puede corroborar según sus criterios de lo que es verdad y lo que puede observar en el momento en que el hablante expresa la oración, dado que, para tener los conceptos de VIENTO, HOJA y LEVANTAR los cuáles implican también a otras muchas creencias aledañas, la intérprete podría basarse en la verdad de esta oración para inferir que el hablante tiene también otras creencias verdaderas, por supuesto, no todas, sino sólo algunas relacionadas con estos conceptos, en principio.

El argumento sintetizado es:

- (HO). Por la naturaleza de la creencia, tener una creencia verdadera supone tener un conjunto de creencias verdaderas (holismo) y que su poseedora es racional.
- (int.) identifico, como intérprete, que un hablante sostiene algo como verdadero.
- PC. Es adecuado interpretar dicha oración sostenida por el hablante como una creencia verdadera. (El PC es verdadero).

Antes de concluir este capítulo, es pertinente señalar que aquí tomé en cuenta únicamente los argumentos para justificar el PC que Davidson ofrece o que se pueden desprender a partir de sus textos, dejando de lado otras justificaciones creativas y quizás fecundas, tales como la de Lepore y Ludwig (2007) que elaboran una justificación del PC como un argumento a la mejor explicación<sup>4</sup>, la de Peter Pagin (2006) quien propone una versión del PC como verdad empírica o la de Kathrin Glüer (2007) que lleva a cabo una formulación del PC como principio necesario a posteriori.

### **Capítulo 3: Las extensiones del PC como herramienta para la evaluación de la racionalidad y una evaluación crítica de las mismas.**

El objetivo de este capítulo es evaluar dos propuestas que existen para extender el PC de un principio que sirve para la interpretación a una herramienta que evalúe la racionalidad del comportamiento humano. Con este fin, desarrollaré los siguientes

---

<sup>4</sup> “Si la suposición de que p se requiere como la mejor teoría en general de acuerdo con la naturaleza de los seres humanos y su lugar en el mundo natural, entonces p; El PC se requiere como la mejor teoría en general de acuerdo con la naturaleza de los seres humanos y su lugar en el mundo natural; por lo tanto, el PC es verdadero”. (Lepore & Ludwig, 2005, 203)

puntos: 3.1) expondré y evaluaré el argumento de la propuesta de extensión de Edward Stein, en 3.2) Llevaré a cabo una evaluación crítica de la posibilidad de la extensión del PC en la versión de Thagard y Nisbett, todo esto para concluir que no es posible una extensión del PC como herramienta para la evaluación de la racionalidad y en 3.3) Elaboraré un breve resumen y conclusiones del capítulo.

### 3.1 La propuesta de Stein sobre la extensión del PC a una herramienta para la evaluación de la racionalidad del comportamiento humano

En su libro, *Without Good Reason*, así como, posteriormente en algunos artículos (Stein, 1997) Edward Stein elabora una evaluación de la racionalidad del comportamiento humano analizando diversos argumentos. Uno de los capítulos de su libro está dedicado específicamente a la posibilidad de extender, lo que él llama, una versión fuerte del PC hacia una herramienta para la evaluación de la racionalidad.

Stein afirma que una estrategia que consista en interpretar una divergencia o error en el razonamiento como un error de ejecución para evitar un error de competencia, sin algún otro tipo de consideración, constituye una “estrategia de inmunización”; es decir, una estrategia *ad hoc* que ayuda a una teoría a resistir frente a la evidencia en contra. *Prima facie*. Ciertamente, el principio de caridad interpreta los errores y las divergencias en el actuar de un agente como errores de ejecución y no de competencia.

Para explicar la racionalidad, Stein toma la Estrategia de la Competencia Racional (ECR) basada en una distinción entre competencia y ejecución, análoga a la que

Chomsky, entre otros autores, proponen para la lingüística y que explicaremos a continuación.

Imaginemos un ejemplo con lo que comúnmente llamamos un *lapsus linguae*. Una persona va a un velorio y saluda a la familia del difunto, los nervios la traicionan y dice: “muchos días de estos”, en lugar de decir algo más adecuado con la situación como: “lo lamento” o “mi sentido pésame”; en general, no pensaríamos en que la persona en cuestión sea un mal hablante de la lengua española, sino que, quizá estaba tan nerviosa que confundió una formalidad que está acostumbrada a decir con otra que dice menos frecuencia, pero que también se encontraba en su archivo de formalidades. En otras palabras, no pensamos que no es competente en la lengua, sino que cometió un error al confundir las frases, es decir, que cometió un error de ejecución. En este caso, los nervios que la persona experimentaba fueron una interferencia que lo llevaron a equivocarse.

Al igual que como lo plantea para el campo de la lingüística, es posible aplicar la distinción competencia-ejecución al razonamiento. Por ejemplo, podemos pensar en algo en que comúnmente no nos equivocaríamos porque lo sabemos muy bien. Supongamos que estoy presionado porque debo avanzar en mi tesis y mientras estoy escribiendo estas líneas, una amiga me pregunta que dónde guardo las botellas de vino y yo le respondo que están en el refrigerador. Si bien no es el único lugar para que las botellas se guarden, tampoco es poco común que alguien meta botellas en el refrigerador, sobre todo, si estuvieran abiertas. Ella va y busca las botellas y no están ahí, me pregunta de nuevo y le respondo que están en la cantina.

Podríamos decir que el estrés de la tesis fue una variable que interfirió en que yo le respondiera correctamente sobre algo que sé muy bien (y sí lo sé muy bien) que las botellas de vino están en la parte de abajo del mueble de la cantina.

Sin embargo, si ella viene por primera vez a mi casa y yo le pido, por favor, que saque una botella de vino, y ella, sin saber absolutamente nada sobre cómo tengo ordenadas mis cosas en mi casa, busca en una puerta de mi librero, no podemos atribuirle el mismo tipo de error que el que yo cometí. Yo sé muy bien en dónde guardo el vino y ella no tiene idea, de modo que ella comete un error de competencia al momento de buscar el vino en un lugar incorrecto, mientras yo cometo un error de ejecución cuando respondo de manera incorrecta a la pregunta sobre dónde guardo el vino.

Por lo tanto, de acuerdo con la noción de racionalidad, basada en la estrategia de competencia racional, los seres humanos somos racionales si los únicos errores que cometemos al razonar son atribuibles a interferencias con nuestros procesos de razonamiento (Stein, 1996), es decir, a algo que obstaculiza nuestra ejecución y no a nuestra habilidad para razonar en sí misma.

### 3.1.1 El papel del PC

Stein hace una revisión de diversas formas en las que Quine utiliza el PC y señala que existe una ambigüedad en su planteamiento, ya que en ocasiones plantea que es menos deseable asumir una mala traducción que una falla en el razonamiento del interlocutor, mientras en otras ocasiones, reconoce que se debe ser caritativo

frente al discurso del otro, a menos que haya evidencia empírica que muestre lo contrario. Ante dicha ambigüedad, Stein utiliza dos versiones distintas del principio de caridad, una débil:

- a) La aserción de una persona debe ser tomada como racional, a menos que haya fuerte evidencia de lo contrario.

Y una fuerte:

- b) Ninguna aserción debe ser tomada como irracional.

Por otra parte, para Stein, Quine no es claro al definir si el PC debe tomarse como una afirmación empírica o conceptual y defiende que no puede ser tomado como una verdad empírica, debido a que, para ello, es necesario tener primero una instancia particular donde se utilice con éxito para llevar a cabo una traducción de la cuál partir como dato empírico, lo cual constituye una petición de principio. Por ello, la única opción es que sea una afirmación conceptual que se funda, sobre todo, en la definición de creencia como algo que sólo lo poseen agentes racionales.

El argumento en el que se sostiene el principio de caridad como una afirmación conceptual, tomado de Daniel Dennett, (Dennett, 1998) de manera sintética, dice que:

- a) Nuestras afirmaciones reflejan nuestras creencias.
- b) Por definición, tener creencias supone la racionalidad de quien las posee.
- c) Por lo tanto, debemos interpretar las afirmaciones de las personas como racionales.

Es decir, lo que hace que un estado mental cuente como una creencia es que está interrelacionado con otros estados mentales (creencias) de manera consistente; por ejemplo, el estado mental  $S$  no puede ser considerado una creencia  $w$  si es posible mantener al mismo tiempo  $w$  y  $\neg w$ . De ahí que podamos afirmar que alguien es racional por el simple hecho de tener creencias. Y si decimos que las afirmaciones de las personas reflejan sus creencias, por lo tanto, se debe interpretar dichas afirmaciones como racionales.

Precisamente, este argumento corresponde con la justificación holística de la naturaleza de las creencias que desarrolla Davidson y que vimos anteriormente.

### 3.1.2 Extensión del PC para la evaluación de la racionalidad del comportamiento

Stein extiende el PC de su utilización en la traducción o la interpretación lingüística a una guía para la interpretación de la competencia de los seres humanos para razonar basándose en lo que, según él, proponen Davidson y Dennett, aunque en los textos que cita no hay como tal una referencia a dicha propuesta por parte de ninguno de los dos, al menos, no explícita, por lo cual, tomaré ésta como una propuesta original suya.

La idea detrás de la posibilidad de esta extensión es que la competencia en el razonamiento es análoga a la traducción o a la interpretación de un enunciado ya que tanto la interpretación del comportamiento de una persona como la de sus afirmaciones requieren que esa persona sea racional, ya que, si no lo fuera, sería imposible atribuirle pensamientos, deseos y creencias. El principio de caridad

supone que la persona que lleva a cabo una afirmación sostiene una creencia y, como lo vimos en el argumento anterior, quien tiene creencias, es racional. De este modo, se puede aplicar el PC para interpretar la competencia en el razonamiento a partir de la conducta de las personas. También, de manera análoga, existe una versión fuerte y una versión débil del principio de caridad aplicado al razonamiento:

- a) débil: a menos que exista evidencia empírica de que evidencie lo contrario, las personas deben ser interpretadas como racionales.
- b) fuerte: las personas nunca deben ser interpretadas como irracionales.

La versión fuerte no niega que haya instancias en las que algunos comportamientos podrían parecer poco o nada racionales, sin embargo, asume que estos casos sólo son errores de ejecución, pero no de competencia.

Stein considera que interpretar la estrategia de la competencia en el razonamiento que consiste en hacer una distinción entre la competencia para razonar y su ejecución concreta, tomada de la distinción que Chomsky hace en lingüística, sin un argumento para hacerlo constituye una estrategia de inmunización y una estrategia ad hoc capaz de negar cualquier evidencia empírica en contra de la teoría. Sin embargo, el principio de caridad provee un argumento para justificar la competencia en el razonamiento, por lo cual, para él, dicha estrategia no es ad hoc.

Veamos un ejemplo de cómo funcionaría (y si funcionaría) la analogía del PC que permite la extensión para Stein.

Siguiendo a Stein, considerando el Experimento de la Conjunción (Tversky & Kahneman, 1983) en donde muchas personas consideran que la probabilidad de  $p$  es mayor que la probabilidad de que  $p$  y  $q$ .

Una interpretación “natural” de los resultados sería que:

Si los seres humanos se caracterizan por una competencia racional que viola sistemáticamente el principio de la conjunción, eso significa que carecen de dicho principio. Por lo tanto, los seres humanos no siempre se comportan de acuerdo con algunos de los principios de racionalidad

Sin embargo, la estrategia de la competencia racional (ECR) más la utilización del PC en su versión extendida permitiría “salvar” la racionalidad humana, interpretando todos los errores en el Experimento de la Conjunción (EC) como errores de ejecución y no errores de competencia, de la siguiente manera:

(IN) La interpretación “natural” de los EC es que los seres humanos violan sistemáticamente los principios normativos del razonamiento. (son irracionales)

(PC) Los seres humanos sólo pueden ser interpretados como seres racionales.

(ECR) Todos los errores de en el Experimento de la Conjunción y otros similares deben ser interpretados como errores de ejecución y no como errores de competencia.

De este modo, la interpretación natural es errónea, los seres humanos en ocasiones fallan en la ejecución, pero eso no hace que deban ser interpretados como irracionales.

Stein defiende que la introducción del PC es un argumento que evita que la introducción ECR sea vista como una estrategia *ad hoc* inmunizante para contrarrestar los resultados empíricos del EC, pero ¿realmente lo hace?

Como sabemos, en general, una estrategia *ad hoc* constituye toda forma de introducción no justificada de un principio o una regla para descartar algún tipo de evidencia que cuente en contra de una teoría. En este caso, si la introducción del PC constituye una estrategia *ad hoc* invalidaría la evidencia empírica que existe en favor de la tesis de que los seres humanos violan sistemáticamente principios lógicos o probabilísticos tal como lo vimos en el EC.

Veamos más a detalle. A continuación, evaluaré si el PC puede verse o no como una hipótesis *ad hoc* desde la propuesta de su extensión como herramienta de evaluación de la racionalidad de Stein.

### 3.1.3 Evaluación del PC dentro de la propuesta de Stein

Siguiendo a Stein, como lo mencioné anteriormente, adoptar la ECR sin un argumento general para hacerlo sería caer en una estrategia de inmunización cuya función sería anular la evidencia empírica en contra de la teoría.

Tal como lo veíamos antes, la estrategia de la introducción del PC, según Stein funciona como argumento para evitar que ECR sea vista como una estrategia de inmunización. Queda entonces claro que el PC salva a la distinción competencia-ejecución racional de ser sólo una forma de blindar a la Tesis de la Racionalidad de la que habla Stein de los resultados empíricos, por ejemplo, del Experimento de la Conjunción. Sin embargo, ¿cómo se justifica la introducción de PC? Para evaluar si está justificado, utilizaré los criterios para evaluar una hipótesis como *ad hoc* que propone Jarrett Leplin en su artículo *The Concept of an "Ad Hoc"* (Leplin, 1975)

*Hypothesis* y los aplicaré al caso de la introducción del PC a la extensión de éste como evaluación de la racionalidad que propone Stein.

El objetivo del trabajo antes citado de Leplin es mostrar, a partir de un análisis de la noción de “*ad hoqueidad*” una defensa de la Hipótesis de la Contracción de Lorentz, y al mismo tiempo, de su adecuación para la teoría de la relatividad, considerada por muchos filósofos, entre ellos Popper, como un ejemplo paradigmático de hipótesis *ad hoc*. En su análisis, Leplin ofrece cinco condiciones para juzgar una hipótesis como *ad hoc*:

Una hipótesis H introducida en la teoría T en respuesta a un resultado experimental E es *ad hoc* si y solo si:

- (1) E es anómalo para T pero no para T como lo complementa H.
- (2) E es evidencia de H, pero:
  - a) no hay resultados experimentales disponibles que no sean E que soporten H,
  - b) H no tiene aplicación para el dominio de T aparte de E,
  - c) H no tiene apoyo teórico independiente.
- (3) No existen suficientes fundamentos para sostener que H es verdadera ni para sostener que H es falso.
- (4) H es consistente con una teoría aceptada y con la propuesta esencial posiciones de T.
- (5) Hay otros problemas además de los que confrontan directamente a E con T que constituyen una razón para decir que están relacionados con E en los siguientes aspectos:
  - a) estos problemas junto con E indican que T no es fundamental;
  - b) ninguno de estos problemas, incluido E, puede resolverse satisfactoriamente, a menos que se elimine esta no fundamentalidad,
  - c) una solución satisfactoria a cualquiera de estos problemas, incluido E, debe contribuir a la solución de los demás. (Leplin, 1975)

Aplicando los criterios de Leplin a la introducción del PC tenemos: que el PC es la hipótesis introducida en la teoría, en este caso, la estrategia de la competencia racional o ECR en respuesta al resultado de los Experimentos de la Conjunción o EC.

1. Los resultados de EC son anómalos para ECR, pero no para ECR si introducimos PC, ya que como Stein mismo acepta, si no introdujéramos PC, entonces, ECR sería *ad hoc* para la Tesis de la Racionalidad.
2. De hecho, los resultados de EC no son evidencia para PC, lo cual nos hace no tener que considerar si quiera los incisos siguientes, aunque lo haremos por mor de la claridad.
  - a) Los resultados experimentales de EC no soportan PC.
  - b) El PC está formulado originalmente para una teoría de la traducción y una teoría de la interpretación radical. Así es que sí tiene otro ámbito de aplicación.
  - c) Si bien, PC tiene varias justificaciones en el ámbito de la Interpretación Radical, no tiene otro ámbito de aplicación para la evaluación de la racionalidad, al menos en esta versión de Stein.
3. Podríamos decir que existen varios argumentos para sostener PC, sin embargo, nuevamente, no están en el campo de la evaluación de la evaluación racionalidad propiamente, aunque sí guardan una relación con ésta.
4. Podemos conceder que PC consistente con una teoría un tanto más aceptada, la Interpretación Radical, y que también es coherente con la posición de ECR.

5. En realidad, el único problema que existe con ECR es que también es una forma adaptada por Stein y otros autores, importada de la lingüística, para analizar la racionalidad del comportamiento humano tales como que su adopción podría exculpar a casi cualquier contra ejemplo empírico resultado de un experimento o a cualquier comportamiento, por alejado de nuestra imaginación que fuera, de constituir una prueba de la irracionalidad del comportamiento humano.

Tomando en cuenta los criterios de Leplin, entonces, es claro que la introducción del PC constituye una hipótesis *ad hoc* para apoyar a ECR en contra de EC a apoyar la Tesis de la Racionalidad y, por lo tanto, tampoco es posible aceptar la propuesta de Stein sobre la extensión del PC para evaluar la racionalidad del comportamiento humano.

Sin embargo, aún sin tomar en cuenta esta evaluación, podemos decir que parece, al menos sospechosa la introducción de PC con el único objetivo de hacer que ECR no se vea como una estrategia de inmunización, lo cual, claramente no ocurre con el papel que juega el PC dentro de la Interpretación Radical, en donde si bien, como vimos, es debatible su indispensabilidad, no está a discusión su funcionamiento, como el propio Stein reconoce, y mucho menos se puede decir que se introduce exclusivamente para justificar a otra hipótesis auxiliar o para “parchar” la teoría. En palabras del mismo Stein:

El principio de caridad está siendo invocado como motivo para no interpretar los resultados de los experimentos sobre el razonamiento experimentos como apoyo a la tesis de la irracionalidad. (Stein, 1996, 132)

En este sentido, la importación de ECR desde el ámbito de la lingüística como la del PC desde el campo de la filosofía del lenguaje, y su posterior unión para tratar de justificar la Tesis de la Racionalidad son, al menos, un tanto forzadas<sup>5</sup>.

Dejaremos la exposición de los argumentos contra la extensión propuesta por Stein hasta aquí, por ahora, ya que un poco más adelante reforzaremos aún más éste argumento de la mano de los argumentos que ofrecen Thagard Nisbett a continuación.

### 3.2 La caridad como método de evaluación en las Ciencias Sociales: La versión de la extensión del PC en Thagard y Nisbett.

En su texto *Rationality and Charity* (1983) Paul Thagard y Richard E. Nisbett se proponen argumentar en favor de la tesis de que la afirmación de que el comportamiento humano diverge de los estándares normativos de racionalidad es una cuestión empírica y que, por lo tanto, la utilización de los PC en las Ciencias Sociales, tales como la Psicología, la Antropología, la Ciencia Política o la Economía, sólo encubre la evidencia sobre las divergencias, y de esta forma, no permite que las ciencias tengan criterios de racionalidad adecuados para los seres

---

<sup>5</sup> Al final, una vez armada su propuesta de extensión del PC como herramienta de evaluación de la racionalidad, Stein mismo la desecha y dice que no presta apoyo a la Tesis de la Racionalidad, aunque, por otras razones que no tienen cabida en este trabajo ya que aquí lo que me propuse es evaluar la posibilidad misma de la extensión, no su funcionamiento.

humanos; para ello, al inicio del texto ofrecen una definición de lo que tomarán como su noción de racionalidad:

Nos interesa principalmente el sentido de "racionalidad" en el que un pensamiento o acción es racional si se ajusta a los mejores estándares normativos disponibles. Por tanto, definimos el comportamiento racional como lo que la gente debería hacer dado un conjunto óptimo de reglas inferenciales.<sup>6</sup> (Thagard & Nisbett, 1983, 251)

Y su definición general de PC como: "Evite interpretar el comportamiento de las personas como violaciones de estándares normativos"<sup>7</sup> (Thagard & Nisbett, 1983, 251), la cual ellos mismos reconocen como ambigua en dos sentidos: el primero, no establece el tipo de estándares normativos que menciona, lo cual para ellos da origen a tres tipos de estándares distintos: 1) principios de caridad para traducción, 2) para las conductas inferenciales y 3) para la toma de decisiones; el segundo sentido de vaguedad es qué tan rigurosa será la forma en que se tome la directiva de evitar atribuir irracionalidad, lo cual da origen a una categorización de PC en cinco grados distintos de mayor a menor grado de rigor:

1. No asumir a priori que las personas son irracionales.
2. No tener ninguna consideración especial en favor de la interpretación de que la gente es irracional.
3. No juzgar a las personas como irracionales, a menos que haya una explicación justificada empíricamente de lo que están haciendo en el momento en que violan los estándares normativos.

---

<sup>6</sup> La traducción es mía: "We are interested primarily in the sense of "rationality" in which a thought or action is rational if it conforms to the best available normative standards. Thus we define rational behavior as what people should do given an optimal set of inferential rules".

<sup>7</sup> La traducción es mía: "Avoid interpreting people as violating normative standards"

4. Interpretar a las personas como irracionales sólo frente a evidencia abrumadora.
5. Nunca interpretar a las personas como irracionales<sup>8</sup>. (Thagard & Nisbett, 1983, 251)

El cruce del tipo de estándar a evaluar según los tres tipos de comportamiento o ámbitos de los que hablan y la clasificación de los cinco principios de caridad según el rigor con el que se apliquen dan lugar a quince combinaciones distintas, de las cuáles los grados 4 y 5, para ellos, no serían deseables.

Posteriormente, los autores analizan cada uno de los tres ámbitos de aplicación del PC. En el caso del primer ámbito, los autores descartan la validez del PC, aplicado a la versión de Quine. A pesar de no ser propiamente una extensión del PC, sino su aplicación a la traducción lo que atacan, dado que sí toca los principios del PC que asenté desde el inicio del trabajo, me permitiré aplicar sus argumentos a la versión de Davidson y dar respuesta a ellos. Posteriormente, expondré su argumento en contra de la extensión del PC para evaluar nuestras conductas inferenciales, y finalmente, sobre el tercer ámbito explicaré brevemente por qué no sería lo óptimo aplicar el PC a la toma de decisiones y me permitiré exponer brevemente una estrategia alterna presente también en Davidson.

### 3.2.1 Los argumentos en contra de la aplicación del PC para la interpretación y su crítica.

El argumento que presentan contra PC en la traducción está fundado en tres contraejemplos ante la postura de Quine con respecto a que frente a un hablante

---

<sup>8</sup> La traducción de las categorizaciones es mía.

que parece estar sosteniendo una contradicción es preferible pensar que llevamos a cabo una mala traducción antes de atribuirle un comportamiento lingüístico irracional. Según su clasificación, esto coloca a Quine en un nivel 3 de rigor del PC; sin embargo, cuando Quine acepta que aun cuando nosotros tengamos como resultado de nuestra traducción de una lengua extranjera una afirmación del tipo “p y no p”, debemos evitar aceptar la contradicción a toda costa, para los autores esto coloca su versión del PC en un nivel 4 o 5.

El caso de Davidson sería análogo o aún más grave, ya que, para los autores, su versión del PC no sólo dice que habría evitar interpretar como inconsistentes las inferencias que llevan a cabo los hablantes, lo cual quiere decir que también se evitarían, *prima facie*, interpretaciones como “p y no p”, contradicciones explícitas, lo que para ellos ya colocaba a Quine en un nivel 4 o 5 como vimos, pero además, en el caso de Davidson, también se deben tomar las creencias del hablante como verdaderas, (también *prima facie*) lo cual muy probablemente lo colocaría en un nivel de rigor 5.

Para elaborar su primer contraejemplo para mostrar la existencia de contradicciones que funcionan de facto en varias lenguas en diferentes culturas, Thagard y Nisbett citan un texto de John Kekes, *A Justification of Rationality*, (Kekes, 1976) en donde se refiere a la discusión del antropólogo británico Evans-Pritchard sobre la religión de la sociedad africana Neur. Kekes afirma que los Neur violan el principio de identidad al afirmar que: “la luz del pantano es espíritu mientras niegan que el

espíritu sea luz del pantano” y justifican que ésta es una traducción correcta al inglés, ya que existe “una amplia comprensión de su lenguaje, su sistema de creencias y sus costumbres”<sup>9</sup> (Thagard & Nisbett, 1983, 253). Concluyen diciendo que:

Comprender creencias ilógicas requiere una batería de técnicas hermenéuticas mucho más potente que la confianza conductista de Quine tiene en los criterios de asentimiento y disentimiento<sup>10</sup>

Es decir, que ni la Traducción radical de Quine y, por ende, ni la Interpretación radical de Davidson están lo suficientemente dotadas de la empiria como los métodos que la antropología tiene para descubrir traducciones acertadas de lenguas extranjeras y que, por ello, no son capaces de dar cuenta de las contradicciones inherentes a los sistemas de creencias.

El segundo contraejemplo proviene del filósofo y antropólogo social británico-checo Ernest Gellner y su libro "Cause and Meaning in the Social Sciences", (Gellner, 1973) en el que defiende que hay muchos grupos sociales donde existen contradicciones culturales que, además tienen funciones sociales importantes como, por ejemplo, el concepto de *agurram* en la cultura de los bereberes de Marruecos. El concepto alude a personas elegidas por dios, pero, de facto, son seleccionadas por miembros de las tribus circundantes que utilizan sus servicios. A

---

<sup>9</sup> *a broad understanding of their language, belief system and customs.*

<sup>10</sup> *Understanding il- logical beliefs requires a much larger battery of hermeneutic techniques than Quine's behavioristic reliance on criteria of assent and dissent.*

pesar de ser contradictorio, dicen, este concepto subsiste por la importancia que tiene en la determinación de comportamientos sociales dentro de la cultura bereber. Con respecto a este segundo contraejemplo, los autores concluyen diciendo: “Debido a que el lenguaje tiene otras funciones además de la de la comunicación de verdades, la traducción no puede ser siempre caritativa” (Thagard & Nisbett, 1983, 254)

Antes de concluir el capítulo, los autores aceptan que los ejemplos que provienen de la antropología siempre son vulnerables frente a las acusaciones sobre que no se ha comprendido del todo a los grupos culturales en cuestión y que, por ende, una investigación más profunda podría revelar que, en realidad no existen contradicciones lógicas en sus sistemas de creencias y que este tipo de cuestiones pueden simplemente no tener una respuesta o no llegar a nada. Por ello, se disponen a ofrecer un contraejemplo que muestre una violación a los principios de la lógica desde lo que ellos llaman “una lógica más cercana a nuestra propia cultura” (Thagard & Nisbett, 1983, 255) Para esto, llaman al ruedo un pasaje de La Ciencia de la Lógica de Hegel citada de la traducción de A.V. Miller, reconocido traductor de varios libros de Hegel al inglés. El texto original que citan en inglés dice: “*Something moves, not because at one moment it is here and at another moment there, but because at one and the same moment it is here and not here, because in this 'here', it at once is and is not.* (Hegel 1969, p. 440).” En la traducción directa del alemán que hace Felix Duque: “Algo se mueve no solamente en cuanto que está aquí en este ahora, y allí en otro ahora, sino en cuanto que es aquí y no aquí en uno y el mismo ahora, en cuanto que está y no está al mismo tiempo en este aquí” (Hegel, 2011, 492) Los autores aclaran que las nociones de negación y de contradicción en

Hegel difieren de la negación y la contradicción veritativo-funcional que utilizan Quine y Davidson, que Hegel acepta las paradojas de Zenón de Elea en contra del movimiento y que, de hecho, este autor busca intencionalmente estas divergencias, retando constantemente a lo largo de toda su obra la forma en la que se han interpretado estos y otros conceptos. Por lo tanto, interpretar a Hegel de forma que se busque evitar la contradicción sería no reconocer la complejidad de su pensamiento. Al final del capítulo, los autores cierran diciendo que aquello que subyace al PC es la concepción de que la función principal del lenguaje es la comunicación; sin embargo, en muchos contextos culturales el lenguaje puede tener muchas otras funciones sociales que sólo se pueden entender si dejamos de pensar en una racionalidad fundada exclusivamente en los principios de la lógica clásica.

Con respecto a los dos primeros supuestos contraejemplos provenientes de la antropología citados por Thagard y Nisbett, el de los Neur y el de los Bereber, podemos decir que, incluso, no es necesario ir muy lejos para encontrar contradicciones en los sistemas de creencias. Quienes profesan el cristianismo aceptan conjuntos de creencias tales como: (1) el padre es dios, (2) el hijo es dios, (3) el espíritu santo es dios, (4) el padre no es el hijo, (5) el padre no es el espíritu santo, (6) el hijo no es el espíritu santo, (7) dios no es el padre, (8) dios no es el hijo, (9) dios no es el espíritu santo y (10) existe solamente un dios. Este conjunto de creencias es sostenido por varios millones de personas en occidente, por lo menos. Sin embargo, creo que éste no sólo no es un tema central dentro de la Interpretación radical, sino que además creo que desvía por completo la atención sobre los problemas que ésta intenta esclarecer.

Antes de responder a las objeciones, me gustaría hacer un excursio rápido, para no salirnos del tema. En el texto de "Deception and Division", Donald Davidson admite, tal como lo hace Quine, que: "Nada de lo que una persona pueda decir o hacer contará como base suficiente para la atribución de una creencia directa y obviamente contradictoria" (Davidson, 1985, 190). De modo que, para él, es posible atribuir a alguien o a un grupo un conjunto no poco problemático de creencias como el que vimos anteriormente, sin que tenga que creer las posibles contradicciones que surjan del mismo. Es decir, es posible que alguien crea (1), (2) y (7) sin que las crea al mismo tiempo, o lo que es lo mismo, que es posible creer en proposiciones que juntas son contradictorias, pero esto no es lo mismo que creer una contradicción: "Es posible creer las dos oraciones (en este caso las tres) sin creer en la conjunción de las dos" (Davidson, 1985, 190). En el caso de los Neur, se puede creer que:

- a) la luz del pantano es el espíritu.
- b) el espíritu no es la luz del pantano.

Sin creer que:

- c) la luz del pantano es el espíritu y el espíritu no es la luz del pantano.

Y esto mismo aplicaría para el caso de los Bereber. Comprendo que esto puede ser controvertible, pero no lo quise dejar pasar, ya que arroja algo de luz sobre lo que es central para el PC en Davidson y aquello que no lo es, pero hasta aquí este excursio por la consistencia de los sistemas de creencias. Vayamos directo a la objeción principal.

Recuperemos los dos primeros capítulos de este trabajo. Traeré dos citas de Davidson puestas anteriormente para elaborar la objeción. Del apartado 1.2.3:

Y, desde luego, el hecho de que una teoría no erija a los hablantes en poseedores universales de verdades no representa una adecuación de la teoría; la meta no persigue el absurdo de hacer desaparecer el disenso y el error. La idea es más bien que solo el acuerdo amplio da el marco dentro del cual pueden interpretarse disputas y errores. La tarea de dar un sentido a las emisiones y conductas de los otros, aún a su conducta más aberrante, nos exige encontrar una gran dosis de razón y de verdad en ellos. (Davidson, 1974a)

Y del apartado 2.1.3:

[...] el acuerdo sobre lo que es abierta y públicamente observable se favorece por encima del acuerdo sobre lo que se oculta, se infiere o es pobremente observado; Las relaciones probatorias deben conservarse cuanto más se acerquen a ser constitutivas del significado. (Davidson 1980, 157).

Dejando de lado la salvedad que los mismos Thagard y Nisbett ponen a las críticas provenientes desde la antropología citada anteriormente, creo que es necesario recordar que el objetivo de la Interpretación radical es la comprensión del lenguaje del hablante o la comunidad de hablantes, tal como lo dice en la primera cita, ofrecer el marco para descubrir acuerdos, desacuerdos, errores e incluso, contradicciones, y que el PC funciona para fijar los significados concediendo racionalidad y la verdad de las creencias del hablante *prima facie* para poder elaborar una interpretación completa de su lenguaje y su sistema de creencias. En ese sentido, como aparece en la segunda cita, el acuerdo sobre lo públicamente observable se favorece por encima de lo que se oculta, como en el caso del tipo de creencias más complejas o elaboradas como si el espíritu es o no es la luz del pantano o sobre si el *agurram* es designado por los dioses o por las tribus bereber, por no hablar sobre todo lo que

se favorecería antes de llegar a interpretar la contradicción hegeliana del párrafo que citan los autores.

Es decir, tratando de ponernos en el papel de intérprete radical davidsoniano, ¿qué harharíamos si nos encontramos con algunos de los contraejemplos anteriores? Que, además, hay que decir que, si bien, son contradicciones presentes en muchísimos grupos (o en todos) culturales en el mundo, constituyen sólo una ínfima parte de las creencias interpretables del sistema de creencias de cada grupo. Si nos encontramos con ellas, muy probablemente tendríamos que dejar su caso como una anomalía dentro de nuestra teoría T y continuar con la interpretación de creencias más básicas o públicamente observables, lo cual, además, no obstaculizaría el que pudiéramos llegar a interpretar miles de creencias más. Una vez con una base suficiente de correspondencia y coherencia de nuestra teoría T y con el conocimiento de que existen contradicciones inherentes a todos los sistemas de creencias y prácticas culturales, podríamos ir más a fondo en ello si fuera nuestro objetivo, aunque hay que recordar que, tal como lo afirman Thagard y Nisbett, es cierto que el objetivo de una teoría de la interpretación radical y dentro de ella del PC no es sustituir el trabajo de un antropólogo, sino optimizar el acuerdo para garantizar la comunicación.

Sobre la pregunta de qué necesitaríamos para interpretar radicalmente la obra de Hegel, considero que, si bien para quienes han estudiado su filosofía y dedican horas de su vida a comprender un párrafo de ésta; esto es una labor titánica, el intentar hacerlo sin conocer a Hegel, sin haber estudiado filosofía y sin saber

alemán, sería, definitivamente, algo muy cercano a lo imposible, pero afortunadamente, también creo que esto, nuevamente, sobrepasa por mucho el objetivo de la teoría.

### 3.2.2 Contra la extensión del PC para las conductas inferenciales

La postura del texto en este apartado es que existe mucha evidencia de que los seres humanos no sólo no venimos equipados con mecanismos lógicos tan refinados, sino que, además, en general, violamos sistemáticamente los principios de la lógica y del razonamiento probabilístico y tenemos varios otros tipos de fallas inferenciales como lo muestran los trabajos de autores como Tversky y Kahneman, (1974) Nisbett y Ross, (1980) Hamill y Wilson (1980), entre otros.

Por otra parte, se encuentran los filósofos como Dennett o Stich, según los autores, que consideran que es necesario asumir cierto grado de racionalidad en los agentes para poder evaluar su competencia inferencial, aún a pesar de los resultados experimentales que niegan que ésta sea una estrategia necesaria. Algunos otros autores como Sober (Sober, 1980) o Jonathan Cohen, (Cohen, 1979 y 1981) quienes presentan un PC reforzado con ECR al igual que lo vimos con el caso de Stein, además importan al ámbito de la racionalidad la distinción que Chomsky desarrolla para la lingüística entre competencia y ejecución racional. Sin embargo, Thagard y Nisbett argumentan que no existe una analogía posible debido a que, si bien se puede defender que el hablante de una lengua tiene por definición un cierto conocimiento de la gramática de su lengua y por ende cierta competencia en ésta,

no existen bases para suponer que la competencia inferencial es innata ni tampoco para defender que es compartida por todos los miembros de un grupo social o una cultura.

Otro argumento que presentan en contra de la analogía de la aplicación de la distinción competencia/ejecución en el ámbito de la racionalidad dice que, ya que es posible mejorar los estándares de razonamiento inductivo y las técnicas inferenciales del mismo, a la luz de los cambios en los métodos y las técnicas inferenciales a lo largo de la historia, mientras que: “No podemos decir que la gramática inglesa haya mejorado en los últimos 400 años” (Thagard & Nisbett, 1983, 259) Hoy existen métodos para evaluar la validez de argumentos mucho más complejos y poderosos que los que proveía la lógica aristotélica, las reglas inductivas utilizadas por los métodos estadísticos y sus aplicaciones a la ciencia han cambiado con el tiempo y han incrementado su eficacia y, en general, nuestros métodos de razonamiento hoy permiten llevar a cabo inferencias mucho más complejas que en el pasado. Lo cual muestra, según estos autores, que la racionalidad no es estática, sino que las inferencias racionales se realizan a partir de las mejores reglas disponibles, y las reglas disponibles mejoran constantemente. Por lo tanto, la competencia racional de las personas tampoco puede ser fija, ya que tiene que ser aprendida y actualizada constantemente, lo cual no ocurre con el lenguaje cuya competencia presenta muy pocos cambios a lo largo del tiempo.

En este sentido, los autores rechazarían la extensión del uso de PC y un PC reforzado con ECR para el caso de las conductas inferenciales, debido a que: 1) El

PC estaría en los niveles de rigor 4 y 5 de su clasificación, lo que en sus propias palabras: “Si nos aferramos a los principios de caridad de nivel (4) o (5), podemos encontrarnos postulando interferencias ad hoc para mantener la presunción de racionalidad” (Thagard y Nisbett, 1983, 258), lo cual presta apoyo a nuestro argumento contra la extensión de Stein, y 2) No es posible la importación del modelo de la distinción competencia/ejecución chomskiana al ámbito de la evaluación de la racionalidad debido, fundamentalmente a que no existen las mismas razones para suponer que hay una analogía relevante entre el conocimiento que tiene de la gramática un grupo social que habla una misma lengua y el que compartan ciertas competencias racionales que les permita hacer las mismas inferencias con el mismo grado de complejidad.

Me gustaría detenerme aquí un poco más para explicar la forma en la que considero que este último punto elaborado por los autores también ofrece apoyo a mis argumentos en contra de sus propios contraejemplos para sostener que el PC es inadecuado también para la traducción y la interpretación radicales.

Considero que la misma razón por la que no es posible importar el modelo de la distinción competencia/ejecución de la lingüística al ámbito de la conducta inferencial, es la que hace que el PC sí sea adecuado para la interpretación radical; es decir, el hecho de que exista una razón para pensar que sí existe una competencia gramatical compartida en un grupo de hablantes, hace pensar que también existe un cierto, aunque sea mínimo grado de racionalidad (al menos dentro de ese ámbito) que se puede presuponer al momento de llevar a cabo la interpretación.

Al parecer, este mismo supuesto o uno similar es el que está debajo de la intuición de Davidson de que las oraciones de creencia están concatenadas unas con otras al modo casi de ser “razones” unas para creer las otras<sup>11</sup>. Una línea por explorar sobre este tema que sólo me gustaría apuntar en este trabajo, pero sobre el que creo que se podría trabajar mucho es la relación que puede tener el PC en Davidson y algunas de sus intuiciones sobre el lenguaje y la racionalidad con las tesis de la teoría de la argumentación pragmática de los lingüistas franceses, Oswald Ducrot y Jean Claude Anscombe que se basa en la idea de que el lenguaje es fundamentalmente argumentativo, aunque para ellos esto no va en la línea de la tradición de la retórica y la dialéctica, sino que quiere decir que una aserción lleva a admitir otras al destinatario de la enunciación. Considero que esto prestaría apoyo a un cierto tipo de racionalidad concatenada inherente al lenguaje mismo que podría complementar la idea de coherencia, uno de los principios del PC en Davidson.

### 3.2.3 Contra la extensión del PC para la toma de decisiones.

Thagard y Nisbett también rechazan la extensión del PC a la teoría de la decisión y, en general, a la ciencia de la decisión que es la economía que está fundada en un modelo normativo de elección racional fuertemente sostenido, el de la teoría de la utilidad esperada (Thagard y Nisbett, 1983, 261), mismo que ha sido defendido a

---

<sup>11</sup> ver el apartado 2.1.2 Principio de Coherencia (PCH) de este trabajo.

ultranza por varios tipos de PC para dejar intactos los principios que lo sostengan, aún a pesar de la evidencia empírica en su contra, lo cual, claramente constituye una estrategia *ad hoc*.

Los autores admiten que los PC utilizados para la decisión tienen que ser aún más fuertes que los requeridos para la inferencia, es decir, de nivel 4 o 5, según su clasificación. Y, además, de constituir una estrategia de inmunización, los rechazan por tres razones fundamentales: la primera, y otra de las razones por lo que los principios deben ser particularmente caritativos en el caso de la decisión, es porque al momento de decidir se ponen en juego las creencias, los deseos y los afectos entre otras cosas, lo cual da pie a que las decisiones tengan de por medio varias interferencias, incluso inconscientes, acepta, que podrían hacerlas pasar como irracionales.

En segundo lugar, en varios experimentos, las decisiones subóptimas en términos monetarios de los sujetos en cuestión se intentan explicar, debido a que dentro de su jerarquía de preferencias se encuentran cosas como la autoestima, u otros aspectos valorados por ellos, por encima del dinero, sin embargo, a pesar de su plausibilidad, no existe evidencia empírica que pruebe esto, y claramente, sin ella, este tipo de objeciones corren el riesgo de constituir hipótesis *ad hoc* en contra de los resultados, y en tercer lugar, requiere demasiado “apoyo caritativo” para la teoría normativa de la decisión racional, que constituye una base importante para varias ciencias sociales, pero que ha sido atacada por varios autores, entre ellos

Kahneman y Tversky, (1974) por no constituir si quiera una base aproximada para los modelos descriptivos de toma de decisión.

### 3.3 Resumen y conclusiones sobre la extensión del PC.

A lo largo de este capítulo, central para el objetivo de este trabajo, revisé la propuesta de extensión del PC de la interpretación a una herramienta para evaluar la racionalidad que Edward Stein construye y después rechaza, que consiste en una estrategia para dar soporte a la Tesis de la Racionalidad en donde el PC se introduce como argumento para justificar la implementación de la distinción competencia/ejecución importada de la lingüística a la racionalidad (ECR) y que esta última no sea sólo vista como una hipótesis *ad hoc*. Sin embargo, Stein no prevé que la introducción misma del PC, además de tener únicamente la función de complementar a ECR sí es *ad hoc*, tal como lo mostré a partir de los criterios elaborados por Jared Leplin y, en ese sentido, no se sostiene la incorporación del PC para evaluar la racionalidad.

Posteriormente, en el texto donde el filósofo Paul Thagard y el psicólogo social Richard E. Nisbett defienden que el que el comportamiento de las personas diverge de los estándares de racionalidad es una cuestión empírica y, por lo tanto, utilizar criterios que escondan o pasen por alto dicha cuestión, como el PC, hacen ineficiente o poco preciso el trabajo de las Ciencias sociales. Su ataque se centra en tres ámbitos de aplicación del PC. La traducción, donde el argumento es que el trabajo de los antropólogos muestra que existen contradicciones en las creencias

de varias culturas que son un contraejemplo a la adscripción de racionalidad con la que opera el PC, a lo que respondí que: 1) pasan por alto que la atribución de racionalidad sólo *prima facie* es para buscar lograr la comprensión y después indagar más, aunque ése no es ya el trabajo de una teoría de la interpretación radical, y 2) el tipo de creencias contradictorias con respecto a ámbitos alejados de lo que puede ser observado directamente, como las creencias religiosas, además de que están presentes en muchas culturas, no obstaculizan la optimización del acuerdo, ya que pueden ser dejadas fácilmente de lado y, aun así, lograr la comprensión de otros múltiples conjuntos que permitirían la comunicación.

Posteriormente, analizan el PC aplicado a las conductas inferenciales de los sujetos, lo cual ya cuenta como una forma de utilización del principio para la evaluación de la racionalidad, y rechazan la utilización del PC en este contexto 1) porque la versión de racionalidad que analizan (la misma que usa Stein), apela a la distinción competencia/ejecución importada de la lingüística, pero que falla como analogía debido a que no existe forma de probar que una comunidad comparta criterios de racionalidad, al menos no necesariamente, tal como sí comparte conocimientos gramaticales, y en segundo lugar, la introducción del PC en este caso, constituye una estrategia de inmunización, un principio *ad hoc* que funciona para descartar la evidencia empírica que existe a favor de que los seres humanos violamos sistemáticamente varios principios normativos en los que se fundan nuestras reglas de racionalidad.

Por último, presenté su argumento en contra de la aplicación del PC a la teoría de la decisión normativa, otro ejemplo de extensión para la evaluación de la racionalidad, principalmente basados en que en la toma de decisiones se involucran elementos afectivos y volitivos que hacen que el nivel de caridad que se debe de utilizar para pasar por alto lo irracional que pueden ser las decisiones considerando esto sea muy alto y corra el riesgo también de constituir simplemente una estrategia *ad hoc*.

En conclusión, aunque Stein sólo construye una extensión del PC para después derrumbarla por no ser adecuada a sus propósitos, su versión de la construcción no se sostiene, al igual que las versiones que presentan Thagard y Nisbett para evaluar tanto el comportamiento inferencial como las decisiones. Desde mi perspectiva, la única aplicación del PC que se sigue sosteniendo, hasta el momento, es la que sirve a la interpretación radical, aunque tampoco ha sido parte de mi tesis hacer una defensa fuerte de ésta, sino sólo analizarla y revisar los argumentos existentes sobre sus posibilidades de extensión para evaluar la racionalidad del comportamiento.

## **4. Conclusiones generales y nuevas posibles rutas de indagación para el PC**

### **4.1 Conclusiones generales**

¿Puede ser extendido el PC como un racionalímetro? Dado lo que hemos visto hasta el momento, el rol del PC en la Interpretación radical es claro: otorga *prima facie* la racionalidad de la intérprete al hablante, no por caridad, sino por estrategia. La idea es fijar momentáneamente la racionalidad y la verdad de las creencias del hablante, o lo que asumimos que el hablante asevera o sostiene como verdadero, para *Ceteris paribus* indagar sobre el significado de éstas y posteriormente, volver para corroborar lo que encontramos y ratificarlo como sus creencias.

Si el aceptamos que el PC cumple con su papel dentro de la teoría de Davidson, hay que hacer énfasis en dos aspectos que, a mi modo de ver son los que más se olvidan al momento de intentar extenderlo, el primero es su carácter de provisional. El PC no atribuye racionalidad a perpetuidad a las creencias del hablante, sino sólo en tanto se construye una teoría lo suficientemente sólida para poder evaluar éstas. Esto es claramente pasado por alto por Stein, Thagard y Nisbett, entre otros autores que parecen sostener que el PC elimina la posibilidad de error o inconsistencia.

La segunda característica sin la cual no puede ser comprendido el PC es el holismo. Además de que no es posible juzgar el PC como una forma de atribución de racionalidad sin tomar en cuenta que la finalidad es hacerlo para construir una teoría del significado, tampoco es posible juzgarlo o tomarlo sin considerar que en la teoría de la interpretación radical se busca llegar a interpretar no una creencia aislada, sino un sistema de creencias, o al menos, la mayor cantidad de creencias posible. La pretensión de Davidson es que la interrelación que existe entre nuestras

creencias solidifique el significado, a partir de solidificar la verdad de las mismas, no olvidemos que él utiliza la teoría del significado de Tarski en la cual el concepto de satisfacción es un término primitivo que se utiliza para llegar a la verdad. Davidson invierte esta relación y toma a la verdad como un concepto primitivo para llegar al significado, apelando a que éste sólo consiste en las condiciones de verdad de una aseveración. En ese sentido, la verdad es una condición de la comunicación y si la comunicación existe, la verdad está garantizada. Esto aunado al requisito de correspondencia PCRR por el cual la intérprete debe tomar en cuenta las circunstancias en el mundo al momento de la emisión del hablante es, en conjunto, lo que garantiza el éxito de la interpretación.

La forma en la que se desprende completamente del contexto el PC y se le despoja de las armas con las cuales cuenta dentro de la teoría de la interpretación radical: la provisionalidad, el holismo, el PCRR y el PCH, lo hacen parecer lo que Wachbroit llama simplemente, “un termómetro de racionalidad” una herramienta que, las más de las veces, es introducida se acepte o no, como en el caso de Stein, como hipótesis *ad hoc* para salvar a los principios normativos constitutivos de una cierta Teoría de la racionalidad del ataque de la evidencia empírica.

Un escollo más, éste no en contra de ninguna extensión de las analizadas en este trabajo, sino uno general que pudiera parecer insalvable para la extensión del PC es que, tal como está planteado no sólo en Davidson, sino también en otros autores como Quine, Wachbroit, Winch, Turner y otros, es el del llamado imperialismo cognitivo, es decir, que finalmente, un PC extendido para evaluar la racionalidad,

como racionalimómetro, los únicos errores de racionalidad que distinguiría serían los nuestros, en palabras de Wachbroit: “puede reconocer los errores que nosotros pudiéramos reconocer y no puede reconocer los que nosotros no pudiéramos reconocer” (Wachbroit, 1987, 44). Sin querer poner en cuestión nuestros criterios de racionalidad basados en los principios de la lógica deductiva, los de la estadística, de la probabilidad, de la inducción, entre otros:

1) la posibilidad de que nosotras y nosotros mismos fallemos al momento de interpretar un enunciado o cometamos un error lógico al momento de enlazar las creencias de un hablante, existe.

2) Tal como Thagard y Nisbett lo dicen, la inferencia racional consiste en llevar a cabo inferencias de acuerdo con las mejores reglas disponibles y las reglas están cambiando y mejorando constantemente (Thagard & Nisbett, 1983, 260), así que nuestra racionalidad no es estática.

3) Sin que quiera decir en ningún sentido que nuestros principios de racionalidad no son válidos, sí es posible que no sean únicos. Cerrar la puerta a otras formas de racionalidad es también cerrar las puertas a otras formas de concebir el mundo.

Considero que este “imperialismo cognitivo” no aplica cuando el PC se utiliza para la interpretación porque, a pesar de ser una imposición de la intérprete hacia el hablante, nuevamente, es sólo una imposición provisional en la búsqueda de optimizar el acuerdo para garantizar la comprensión y en pos de la comunicación. Sin embargo, creo que esta discusión sale del objetivo de este trabajo.

Por último, quiero decir una razón más, también general, que va más allá de sólo mostrar por qué el PC no puede ser utilizado como herramienta para evaluar la

racionalidad, sino que, además, en la versión de Davidson está presupuesto ya en cada prueba. Davidson mismo se declara escéptico sobre la eficacia de las pruebas empíricas de las teorías de racionalidad:

Soy profundamente escéptico sobre la posibilidad de pruebas experimentales significativas de teorías de la racionalidad. Esto no significa que tales teorías, o las consideraciones que se encuentran detrás de ellas, no tengan aplicación. Por el contrario, pienso en estas teorías como intentos de iluminar un aspecto esencial de los conceptos de creencia, deseo, intención y significado. Un criterio del que una teoría sobre estos conceptos debe dar cuenta es éste: debe mostrar cómo es posible para una persona ("el experimentador ") llegar a comprender a otra (" el sujeto ") (Davidson, 1985)

En el sentido de racionalidad del que Davidson habla, hacer una prueba de racionalidad cualquiera presupone lo que quiere probar. El PC en la interpretación radical asume la racionalidad para encontrar el significado, la creencia, las preferencias para, incluso, llegar a la interpretación de la acción. Sin embargo, la principal objeción con respecto a llevar a cabo pruebas sobre comportamientos racionales es que, si Davidson tiene razón, se presupone la comprensión lingüística en la mera acción de llevar a cabo la prueba. En realidad, lo que se quiere hacer es medir el nivel de racionalidad que se tiene con respecto a los estándares normativos de racionalidad de la lógica y de la teoría de la probabilidad, pero para ello se asume que la comprensión de las oraciones escritas, las intenciones de los investigadores e incluso el conjunto de acciones que se llevan a cabo previas a elaborar las pruebas se sobreentiende. En este sentido, antes de la prueba, el PC ya fue puesto en marcha. En realidad, si Davidson tiene razón, la racionalidad es algo que tenemos ya siempre que presuponemos interpretabilidad, es decir, damos por supuesto que

nuestro sujeto ya comprendió nuestros significados y que lo hace de la misma forma como nosotros y nosotras intentamos plasmarlos.

#### 4.2 Otras rutas de indagación para el PC dentro del análisis de la racionalidad.

Hasta donde mostré, sin algún otro tipo de restricciones o, el PC no puede ser utilizado como un método para evaluar la racionalidad del comportamiento humano.

Robert Wachbroit concuerda con ello y, en su artículo, *Theories of Rationality and Principles of Charity* abre una puerta distinta al PC, más allá de sólo una herramienta de racionalidad o, lo que él llama, un “termómetro de la racionalidad”. Wachbroit considera que el PC, más allá de una herramienta, puede ser considerado como una característica fundamental que sea la base de una Teoría de la racionalidad que resulte conveniente para la evaluación de teorías científicas a los filósofos e historiadores de la ciencia.

Para Wachbroit una teoría de la racionalidad científica tiene fundamentalmente tres tareas:

- 1) Responder al escéptico, en el sentido en que una teoría de la racionalidad es una teoría de la justificación en el sentido análogo al de la epistemología.
- 2) Proveer un análisis sobre en qué consisten los juicios de racionalidad, si se entienden únicamente en términos de la lógica deductiva o también apelan a consideraciones sobre probabilidad, por ejemplo.
- 3) Establecer un punto de partida desde el cual se puedan dirimir las controversias sobre racionalidad, en el sentido de si se consideran sólo

aspectos internos a las disputas entre teorías científicas o también aspectos externos históricos o sociológicos.

Para él, tradicionalmente, una teoría de la racionalidad lo que hace es establecer criterios o reglas que especifican las condiciones bajo las cuáles algo constituye una buena razón para creer en algo más. De este modo, toda teoría científica debe estar sustentada en una teoría de la racionalidad bajo la cual se especifiquen los criterios de lo que se confirma como un juicio verdadero para dicha teoría y lo que no. Esta versión tradicional, que denomina la Teoría de la racionalidad basada en criterios, sin embargo, tiene la dificultad de no poder especificar el conjunto de todas las reglas para establecer de manera clara los criterios, de modo que el juicio de los científicos es considerado como parte de la evidencia para la teoría, lo cual, de ninguna manera es ni ha sido un problema que cuente en contra de esta concepción; sin embargo, sí abre la puerta para el escepticismo y el relativismo, lo cual para Wachbroit es una oportunidad para una Teoría de la racionalidad basada en el PC.

Por un lado, dado que una Teoría de la racionalidad basada en el PC no pretendería establecer el total del conjunto de reglas de racionalidad en el que se funda una teoría, ya que no distingue entre lo que es racional para nosotras y lo que realmente es racional, por lo cual, no abre la puerta al escéptico y, por otro lado, tampoco distingue entre nuestra determinación de lo que debe ser verdadero y lo que es verdadero apelando a un criterio externo, por lo que no analiza los juicios de racionalidad como requiere el segundo criterio para una teoría de la racionalidad,

pero tampoco es susceptible del relativismo al modo en que una Teoría de la racionalidad basada en criterios, lo es.

Por otro lado, revisando los criterios generales para toda teoría, una Teoría de la racionalidad basada en el PC: 1) no establece una teoría de la justificación porque no lo requiere, pero tampoco es susceptible de escepticismo y 2) tampoco provee un análisis sobre los juicios de racionalidad porque no lo requiere, aunque sí 3) ofrece una forma de adscripción de racionalidad, por ejemplo, cuando se trata de revisar la historia de la ciencia. “Una historia de la ciencia informada por una Teoría de la racionalidad basada en el PC da como resultado un tipo de historia Whig”. Si bien es cierto que comúnmente se considera una historia Whig como una forma de hacer historia que justifica que lee el pasado a la luz de los logros del presente justificando una idea ingenua de progreso, Wachbroit considera que existen otros tipos de versiones Whig no ingenuas que son útiles a los historiadores o a los filósofos de la ciencia.

Otra propuesta contemporánea que incorpora el PC como parte de sus principios de trabajo es el Análisis Racional: un programa de investigación empírica desarrollado inicialmente por el psicólogo John Anderson que tiene como objetivo explicar por qué el sistema cognitivo es adaptativo respecto a sus propósitos y a la estructura del medio. En palabras de Mike Oaksford, uno de sus principales exponentes: “Esta metodología provee de un marco teórico para explicar el nexo que existe entre los principios formales de racionalidad y el éxito práctico de la racionalidad cotidiana, no sólo en psicología, sino en el estudio del comportamiento”

(Chater y Oaksford, 2000, 106). En este sentido, los principios formales de racionalidad sirven para establecer la ruta óptima para llegar a las metas que tiene un sistema cognitivo en un medio determinado. La idea es que en la cotidianidad dicho sistema tendrá éxito en la medida en que se aproxime a la ruta establecida por dichos principios. El Análisis Racional se aplica con éxito en psicología, economía y sirve también para evaluar el comportamiento animal, así como otras áreas donde el objetivo sea optimizar una meta.

Tal como lo defiende Oaksford, el programa de investigación utiliza el PC de manera análoga a como ocurre en la Interpretación radical de Davidson para interpretar el comportamiento de los participantes en los experimentos, de acuerdo con las reglas de la teoría normativa que mejor lo describan, es decir, las que den un mejor sentido racional a dicho comportamiento. En este sentido, la analogía de la utilización se sostiene, ya que lo que hace el Análisis racional no es utilizar el PC para evaluar la racionalidad de los participantes, sino interpretarlos según la regla que más convenga, al igual que en la interpretación radical para: “tratar de dar el mejor sentido a lo que nos han dicho”, es decir, buscando la comprensión.

Tal como lo mencioné anteriormente y como lo retoma también Oaksford, el programa de Davidson no culmina con la comprensión del hablante por parte de la intérprete en la interpretación radical, sino que va más allá. Dado que Davidson no comparte del todo el conductismo de Quine, no es suficiente sólo con interpretar el comportamiento públicamente observable de un hablante para conocer el significado de sus aserciones y sus creencias, sino que el programa completo, la interpretación del lenguaje, el pensamiento y la acción de un hablante requiere que

conozcamos también sus deseos e, incluso la fuerza con la que cree en una proposición, su grado de creencia o, lo que es lo mismo en la versión bayesiana de la teoría de la decisión, sus preferencias y las probabilidades subjetivas que atribuye a los posibles sucesos. Tal como lo dijimos Davidson toma para encontrar estos dos desconocidos el método de apuestas elaborado por Frank Ramsey y los axiomas de Savage para mostrar como cuando las preferencias de una persona cumplen con ciertos requisitos hay probabilidades y utilidades que garantizan que sus preferencias maximicen la utilidad esperada. De este modo, los deseos o las preferencias y los grados de creencia o las probabilidades de un agente se pueden inferir por medio del método de las apuestas, y las creencias y los significados por medio del PC en la interpretación radical.

En resumen, tal como lo vimos, Robert Wachbroit abre la puerta a la posibilidad de elaborar una teoría de la racionalidad científica basada en el PC que si bien no es capaz de proveernos un conjunto de reglas a partir de las cuáles elaboremos criterios para establecer juicios sobre racionalidad, sí podría ser muy útil para adscribir racionalidad cuando se especifiquen las circunstancias o el contexto en el cual se realiza un juicio. El caso concreto, para él, es la historia de la ciencia y la filosofía de la ciencia.

Por otra parte, el programa del Análisis Racional utiliza el PC como criterio para discernir en los experimentos entre varias teorías normativas y reglas, cuál es la que hace más racional el comportamiento de un individuo, desde luego, a ojos del científico que la aplica. Es decir, comparte el objetivo del PC dentro de la interpretación radical que presupone la racionalidad para llegar a la comprensión

del lenguaje de un hablante, y en el caso del Análisis Racional, del comportamiento de un participante.

Por último, otra vía que ha dado buenos frutos para la utilización del PC, también análoga a la forma en la que se utiliza en la Interpretación Radical y en el Análisis Racional, es la argumentación. En donde, dado un cierto fragmento de discurso donde asumimos que existe un argumento, suponemos la racionalidad *prima facie* e intentamos reconstruir la versión más sólida de un argumento, en la medida de lo posible, para evitar caer en una falacia de hombre de paja y para poder evaluar el argumento en su mejor versión dado el principio griego de que quien puede lo más, puede lo menos y no viceversa.

## Bibliografía y Referencias

Anscombe, J. and Ducrot, O., (1994). *La Argumentación En La Lengua*. Madrid: Gredos.

Cabrera, I. (1999). Analítico y sintético, *a priori* y *a posteriori*. En Villoro, L (Ed.) (1999). *El Conocimiento*. Trotta: Madrid, pp. 135-164.

Cabrera, I. (comp.). (1999). *Argumentos trascendentales*. IIF-UNAM: México, pp. 7-31.

Cohen, L. J. (1979), "On the Psychology of Prediction: Whose is the Fallacy?", *Cognition* 7: 385-407.266 PAUL THAGARD AND RICHARD E. NISBETT

- (1981), "Can Human Irrationality be Experimentally Demonstrated?", *Behavioral and Brain Sciences* 4: 317-331.

Chater, N. and Oaksford, M., (2000). *The Rational Analysis of Mind and Behavior*. *Synthese*, 122, pp.93-131.

Donald, D., (1967). *Verdad y significado*. En: Davidson, D. (2001). *De la verdad y de la interpretación*, 6th ed. Barcelona: Gedisa, pp.39-57.

- (1970), *Sucesos mentales*. En Davidson, D. (1995). *Ensayos sobre acciones y sucesos*. México: UNAM/Crítica. 263-285.

- (1973). *Interpretación radical*. En: Davidson, D. (2001), *De la verdad y de la interpretación*, 6th ed. Barcelona: Gedisa, pp.137-151.

- (1974a). *La creencia y el fundamento del significado*. En: Davidson, D., *De la verdad y de la interpretación*, 6th ed. Barcelona: Gedisa, pp.151-164.

- (1974b). *De la idea misma de un esquema conceptual*. En: Davidson, D. (2001), *De la verdad y de la interpretación*, 6th ed. Barcelona: Gedisa, pp.189-204.

- (1974c) *La psicología como filosofía*. En Davidson, D. (1995). *Ensayos sobre acciones y sucesos*. México: UNAM/Crítica. 289-302.

- (1975). *Pensamiento y habla*. En: Davidson, D. (2001), *De la verdad y de la interpretación*, 6th ed. Barcelona: Gedisa, pp.164-179.

- (1980). *A Unified Theory of Thought, Meaning and Action*. En: *Problems of Rationality*. Oxford: Clarendon Press 2004, pp. 151–166.

- (1982a). *Comunicación y convención*. En: Davidson, D. (2001), ed., *De la verdad y de la interpretación*, 6th ed. Barcelona: Gedisa, pp.189-204.

- (1982b). Paradoxes of Irrationality. En: *Problems of Rationality*. Oxford: Clarendon Press 2004, pp. 169-189.
- (1983). Una teoría coherentista de la verdad y el conocimiento. En: Davidson, D., (2003). *Subjetivo, Intersubjetivo, Objetivo*. Madrid: Ediciones Cátedra.
- (1985a) A New Basis for Decision Theory, *Theory and Decision* 18: 87–98.
- (1985b). Paradoxes of Irrationality. En: *Problems of Rationality*. Oxford: Clarendon Press 2004, pp. 189–199.
- (1986). Deception and Division. En: *Problems of Rationality*. Oxford: Clarendon Press 2004, pp. 199–213.
- (1991). Tres variedades de conocimiento. En: Davidson, D., (2003). *Subjetivo, Intersubjetivo, Objetivo*. Madrid: Ediciones Cátedra
- Dennett, D. (1988). La actitud intencional. Barcelona: Gedisa.
- Gauker, C., (1986). The Principle of Charity. *Synthese*, 69, pp.1-25.
- Gellner, E. (1973). Cause and Meaning in the Social Sciences. Nueva York: Routledge.
- Glüer, K., (2007). The Status of Charity I: Conceptual Truth or A Posteriori Necessity? *International Journal of Philosophical Studies*, 14(3), pp.337-359.
- (2011). *Donald Davidson A Short Introduction*. Nueva York: Oxford University Press.
- Goldberg, N., (2004). The Principle of Charity. *Dialogue*, 43(4), pp.671-683.
- Grünbaum, A., (1976). Ad Hoc Auxiliary Hypotheses and Falsificationism. *The British Journal for the Philosophy of Science*, 1976(4), pp.329-362.
- Hansberg, O., (1988). Interpretación radical en Donald Davidson. *Ergo*, 2(4), pp.41-49.
- Hegel, G. y Duque, F., (2011). *Ciencia De La Lógica*. Madrid: UAM.
- Henderson, D., 1987. The Principle of Charity and the Problem of Irrationality. *Synthese*, 73, pp.225-252.
- Hunt, C., (2012). On Ad Hoc Hypotheses. *Philosophy of Science*, 79(1), pp.1-14.
- Kekes, J. (1976). A Justification of Rationality. Nueva York: Albany State University of New York Press.
- Leplin, J., (1975). The Concept of an Ad Hoc Hypothesis. *Studies in History and Philosophy of Science*, 5, pp.309-345.

Lepore, E. and Ludwig, K., (2007). *Donald Davidson: Meaning, Truth, Language And Reality*. 2nd ed. Great Britain: Oxford University Press.

- (2013), *A Companion to Donald Davidson*. Oxford: Wiley-Blackwell.

Mele, A. and Rawling, P., (2004). *The Oxford Handbook Of Rationality*. Oxford: Oxford University Press.

Mulhall, S., (1987). Davidson on Interpretation and Understanding. *The Philosophical Quarterly*, 37(148), pp.319-322.

Nisbett, R. and Ross, L. (1980), *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs: Prentice Hall.

Oaksford, M., (2014). Normativity, interpretation, and Bayesian models. *Frontiers in Psychology*, 5, pp.1-5.

Pagin, P., (2006). The Status of Charity II. Charity, Probability, and Simplicity. *International Journal of Philosophical Studies*, 14(3), pp.361-38.

- (2013). Radical Interpretation and the Principle of Charity. En: Ernie Lepore y Kirk Ludwig (ed.), (2013), *A Companion to Donald Davidson* (pp. 225-246). Oxford: Wiley-Blackwell

Stein, E., (1996). *Without Good Reason*. New York: Clarendon Press.

- (1997). Can We Be Justified in Believing that Humans Are Irrational? *Philosophy and Phenomenological Research*, 57(3), pp.545-565.

Stich, S. (1993). *The Fragmentation of Reason*.

Sober, E. (1978), "Psychologism", *Journal for the Theory of Social Behavior* 8: 165-191.

-(1980), "The Evolution of Rationality", paper read to Society for Philosophy and Psychology.

Thagard, P. and Nisbett, R. (1983). Rationality and Charity. *Philosophy of Science*, 50(2), pp.250-267.

Tversky, A. and Kahneman, D. (1974), "Judgment Under Uncertainty: Heuristics and Biases", *Science* 185: 1124-1131.

-(1983). Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychological Review*, 99, pp.293-315.

Wachbroit, R., (2020). Theories of Rationality and Principles of Charity. *The British Journal for the Philosophy of Science*, 38(1), pp.35-47.